

University of Alberta

**Towards the Complete Backbone Chemical Shift
Assignments of HAV 3C Protease**

by

Michael Scott Watson ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial
fulfillment of the requirements for the degree of

Master of Science

in

Pharmaceutical Sciences

**Faculty of
Pharmacy and Pharmaceutical Sciences
Edmonton, Alberta
Spring, 2000**



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-60191-9

Canada

University of Alberta

Library Release Form

Name of Author: Michael Scott Watson

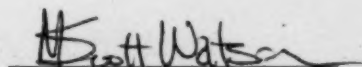
Title of Thesis: Towards the Complete Backbone Chemical Shift
Assignments of HAV 3C Protease

Degree: Master of Science

Year this Degree Granted: 2000

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.



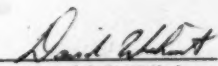
Faculty of Pharmacy and
Pharmaceutical Sciences
University of Alberta
Edmonton, Alberta
T6G 2C2

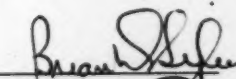
Date: Jan. 27, 2000

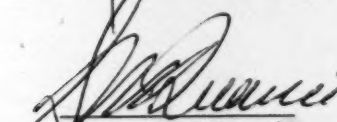
University of Alberta

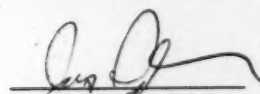
Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "Towards the Complete Backbone Chemical Shift Assignments of HAV 3C Protease" by Michael Scott Watson in partial fulfillment of the requirements for the degree of Master of Science in Pharmaceutical Sciences.


Dr. David S. Wishart


Dr. Brian D. Sykes


Dr. Steve McQuarrie


Dr. Jim Diakur

Date: Jan. 27, 2000

Abstract

The medically significant Hepatitis A Virus (HAV) 3C protease processes the polyprotein produced by HAV. In this thesis, we describe our efforts at characterizing HAV 3C protease using various approaches including analytical ultracentrifugation, isotopic labeling and NMR spectroscopy with the specific goal of completing the backbone ^1H , ^{13}C , and ^{15}N assignments of the HAV 3C protease. Chapter 2 discusses the analytical ultracentrifugation, ^{15}N T_2 -relaxation, and ^1H linewidth studies that identified sample conditions where HAV 3C protease exists as a monomer at suitable concentrations for NMR spectroscopy. Chapter 3 describes the strategies that facilitated the expression and purification of large quantities (>30 mg/l) of uniformly (^{15}N , $^{13}\text{C}/^{15}\text{N}$, and $^2\text{H}/^{13}\text{C}/^{15}\text{N}$) and ^{15}N -residue specific labeled HAV 3C protease from *E. coli*. Finally, chapter 4 discusses the heteronuclear multidimensional NMR experiments performed that permitted ^1H , ^{13}C and ^{15}N chemical shift assignments for $\sim 85\%$ of the backbone.

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor, Dr. David Wishart, for his encouragement, enthusiasm and help. Without the first impressionable discussion I had with him in September of 1996 and his encouragement throughout this program, I would never have been able to accomplish this work.

I would also wish to express my whole-hearted thanks to all my colleagues and fellow students that I have had the pleasure of associating with over the past three years. In particular, I would like to thank Alan Gibbs, Don Husereau, Gary Van Domselaar, Ashenafi Abera and Mesfin Fanta who have all been tremendous friends, providing scientific assistance, many intellectual conversations and a continual source of emotional support. I would also like to thank Ashenafi Abera, Scott Ellerbeck, Hassan Monzavi, YunJun Wang, Alex Nip, Yamini Rammamoortha and Trent Bjorndahl for all their assistance on the HAV 3C Project.

I am very grateful to Dr. Brian Sykes, Dr. Carolyn Slupsky, Dr. Leo Spyropoulos and Dr. Stephane Gagne for their spectroscopic help and their valuable teachings over the past three years. I am also grateful to Dr. Bruce Malcolm who supplied the E. coli expression system with which we initiated this work and with whom I have had many helpful, insightful discussions as well. In addition, I would like to thank Dr. M.N.G. James, Dr. John Vederas and members of their respective groups for their valued contributions to this study. For technical assistance, I would like to Les Hicks (Sedimentation Analysis), Robert Boyko and Lee Willard (Computational) and all the members of PENCE. I would also like to express my appreciation to members of the Faculty of Pharmacy, Dr. Steve McQuarrie, Dr. John Bachynsky, Dr. Jim Diakur and Dr. Brian Amsden, for their friendship, advice and assistance throughout my studies.

I particularly wish to express my deep gratitude to my entire family, Anthony, Laurie, Hugh, Shannon and Dan for their love, sacrifice and encouragement over the past three years.

Finally, I would like to acknowledge Bristol Myers Squibb and PENCE for financial and technical support throughout my graduate career.

Table of Contents

	Page
Chapter 1: Introduction.....	1
1.1 NMR Spectroscopy.....	4
1.2 Sequence-Specific Assignments of Larger Proteins.....	19
1.3 Hepatitis A Virus 3C Protease.....	27
Chapter 2: Determining solution conditions that minimize aggregation of HAV 3C protease.....	37
2.1 Introduction.....	37
2.2 Materials and Methods.....	40
2.3 Results and Discussion.....	52
2.4 Conclusion.....	67
Chapter 3: ^2H, ^{13}C and ^{15}N isotopic labelling of HAV 3C protease.....	68
3.1 Introduction.....	68
3.2 Materials and Methods.....	70
3.3 Results and Discussion.....	77
3.4 Conclusion.....	84
Chapter 4: NMR experiments and Backbone Assignments.....	85
4.1 Introduction.....	85
4.2 Heteronuclear NMR experiments used to obtain sequential backbone assignments for HAV 3C protease.....	87
4.3 Materials and Methods.....	100
4.4 Results and Discussion.....	107
4.5 Conclusion.....	118

4.2
Chapter 5: General Discussion and Conclusion.....120

Appendix A: Processing NMR spectra with NMRPIPE software.....126

References.....133

List of Tables

	Page
Table 1.1	2
Table 1.2.....	16
Table 1.3.....	23
Table 1.4.....	28
Table 2.1.....	45
Table 2.2.....	50
Table 2.3.....	65
Table 2.4.....	66
Table 3.1.....	70
Table 3.2.....	71
Table 3.3.....	76
Table 3.4.....	78
Table 3.5.....	83
Table 4.1.....	101
Table 4.2.....	109
Table 4.3.....	116
Table 4.4.....	118

List of Figures

	Page
Figure 1.1.....	6
Figure 1.2.....	8
Figure 1.3.....	11
Figure 1.4.....	17
Figure 1.5.....	18
Figure 1.6.....	21
Figure 1.7.....	26
Figure 1.8.....	30
Figure 1.9.....	31
Figure 1.10.....	34
Figure 1.11.....	35
Figure 1.12.....	36
Figure 2.1.....	39
Figure 2.2.....	42
Figure 2.3.....	47
Figure 2.4.....	48
Figure 2.5.....	49
Figure 2.6.....	55
Figure 2.7.....	57
Figure 2.8.....	60
Figure 2.9.....	62
Figure 2.10.....	64
Figure 3.1.....	74
Figure 3.2.....	81
Figure 3.3.....	82
Figure 4.1.....	93
Figure 4.2.....	94
Figure 4.3.....	95
Figure 4.4.....	96
Figure 4.5.....	97

Figure 4.6.....	98
Figure 4.7.....	99
Figure 4.8.....	104
Figure 4.9.....	105
Figure 4.10.....	106
Figure 4.11.....	115

List of Abbreviations

Amino acids:

A (Ala)	=	L-alanine
C (Cys)	=	L-cysteine
D (Asp)	=	L-aspartic acid
E (Glu)	=	L-glutamic acid
F (Phe)	=	L-phenylalanine
G (Gly)	=	L-glycine
H (His)	=	L-histidine
I (Ile)	=	L-isoleucine
K (Lys)	=	L-lysine
L (Leu)	=	L-leucine
M (Met)	=	L-methionine
N (Asn)	=	L-asparagine
P (Pro)	=	L-proline
Q (Gln)	=	L-glutamine
R (Arg)	=	L-arginine
S (Ser)	=	L-serine
T (Thr)	=	L-threonine
V (Val)	=	L-valine
W (Trp)	=	L-tryptophan
Y (Tyr)	=	L-tyrosine

2D	=	two dimensional
3D	=	three dimensional
CHAPS	=	3-[(3-cholamidopropyl)-dimethyl-ammonio]-1-propanesulfonate
COSY	=	correlation spectroscopy
CSA	=	chemical shift anisotropy
CT	=	constant time
DD	=	dipole-dipole
DLVO	=	Deryagin-Landau-Verwey-Overbeek theory of colloids
DTT	=	dithiothreitol
E. coli	=	Escherichia coli
EDTA	=	ethylenediaminetetraacetic acid
FID	=	free induction decay
HAV	=	hepatitis A virus

HMQC	=	heteronuclear multiple quantum correlation spectroscopy
HSQC	=	heteronuclear single quantum correlation spectroscopy
INEPT	=	insensitive nuclei enhanced polarization transfer
IPTG	=	isopropyl β -D thiogalactopyranoside
J	=	coupling constant
kDa	=	kilodalton
MM294	=	prototrophic strain of E. coli containing pHAV-3CEX
ncyc	=	number of cycles
NMR	=	nuclear magnetic resonance
NOE	=	nuclear Overhauser effect
NOESY	=	nuclear Overhauser effect spectroscopy
OD ₂₈₀	=	optical density at 280 nm
OD ₆₀₀	=	optical density at 600 nm
PDB	=	Protein Data Bank
PW_SHPSS	=	cosine-modulated 180° ¹ H pulse
PWN	=	90° ¹⁵ N pulse
SAR by NMR	=	structure activity relationships by nuclear magnetic resonance
SDS-PAGE	=	sodium dodecyl sulfate polyacrylamide gel electrophoresis
T ₁	=	longitudinal or spin-lattice relaxation rate
T ₂	=	transverse or spin-spin relaxation rate
TFE	=	trifluoroethanol
TOCSY	=	total correlation spectroscopy
γ	=	gyromagnetic constant
$\Delta\nu_{1/2}$	=	half-height linewidth
τ_c	=	correlation time

Chapter 1

Introduction

Over the past two decades, advances in structural biology have radically transformed the ways in which drugs are designed and discovered. Today, rather than treating protein drug targets as simple "black boxes", scientists are now treating them as complex molecular entities that possess well-defined three dimensional structure and specific sites that can be activated or deactivated by small molecule ligands. Innovations in structure determination methods and computer-aided molecular rendering techniques have led to the development of 'structure based drug design' or 'rational drug design'. These are drug development techniques that use the structure of a protein target as an essential template in the design or construction of protein-specific small molecule ligands (Goody, 1995).

One class of proteins that is particularly applicable to rational drug design is the proteases. Proteases are enzymes that cut or cleave peptides and proteins at specific sites (Voet and Voet, 1995). They are critical to many vital processes including digestion, protein turnover, immune recognition, and protein processing. Proteases have been classified into four families: cysteine proteases, metalloproteases, aspartic proteases, and serine proteases based on their catalytically active residues. It is known that a number of proteases are involved in a variety of human diseases (Table 1.1). Consequently, many proteases could serve as excellent targets for the development of inhibitors and new therapeutic agents.

Interestingly, the first "rationally designed" drug (developed by researchers at Squibb) inhibited a protease known as angiotensin converting enzyme (ACE) (Rubin et al., 1978). This drug, named captopril, was the first of a number of impressive ACE-based antihypertensive medications that have appeared over the

last 20 years. In fact ACE inhibitors are by far the most significant rationally designed drugs to have reached the market. Since 1975, there have been several other successful examples of rational designed drugs including the HIV protease inhibitors (DesJarlais et al., 1990; Olson and Goodsell, 1998) and anti-tumor agents such as capecitabine (Verweij, 1999).

Table 1.1 Protease Families

Family	Representative Enzyme	Disease
Serine Proteases	Neutrophil elastase	Pulmonary emphysema
	Thrombin	Platelet dependent thrombus formation
Metalloproteases	Angiotensin converting enzyme (ACE)	Hypertension
	Stromelysin	Inflammatory disorders
Cysteine Proteases	HAV 3c protease	Hepatitis A
	Cathepsin B	Tumor formation
	Calpain	Stroke
Aspartic Proteases	Renin	Hypertension

The rational design of any given drug depends on a precise knowledge of the three-dimensional (3D) structure of the protein target. To date, the most prominent technique for 3D structure determination has been X-ray crystallography (Blundell and Johnson, 1976). More recently, structure-determination techniques based on nuclear magnetic resonance (NMR) spectroscopy have emerged as a powerful alternative to X-ray crystallography (Wider and Wuthrich, 1999). NMR spectroscopy is often complementary to X-ray crystallography, with one method providing structural or dynamic information that is not as easily obtained with the other. This was demonstrated recently as information derived from both techniques was used in the structural refinement of

bovine pancreatic trypsin inhibitor (BPTI) (Schiffer et al, 1994). Using NMR spectroscopy to derive solution structures of proteins can also provide a check for the correctness of a previously determined crystal structure (and vice versa). Indeed, discrepancies detected between NMR and X-ray structures have increased awareness about the possible mistakes that can be made using either technique (Wagner, 1993). What makes NMR spectroscopy particularly attractive is that the protein of interest is normally studied in solution as opposed to the solid-state that is required for X-ray crystallography. For these reasons structural analysis by NMR is clearly a worthwhile endeavor, even in cases where an X-ray structure of the same protein has already been solved.

This dissertation describes the use of NMR spectroscopy to characterize a large, medically significant protein that could serve as a suitable target for rational drug design. Specifically, it centers on the 217 residue 3C protease from the Hepatitis A virus. This particular 3C protease is responsible for processing the polyprotein produced by the translation of viral mRNA and therefore is essential to the propagation of this virus. Consequently, it serves as an excellent "model" protein target for the rational design of an anti-viral drug.

Although the HAV 3C protease structure has been previously determined by X-ray crystallography (Allaire et al., 1994), we have chosen to characterize this protein via NMR spectroscopy. The intention is to use this technique to help screen for putative inhibitors via a process called "SAR (Structure Activity Relationships) by NMR" (Shuker et al., 1996). More specifically, this thesis describes the efforts to characterize and "condition" HAV 3C protease which, ultimately resulted in the near-complete backbone assignments of this protein. These assignments will serve as a critical first step for future SAR by NMR studies. In particular, chapter 2 describes the techniques that were used to determine sample conditions at which HAV 3C protease existed as a monomer – a prerequisite for the collection of "good" NMR spectra. Chapter 3 describes the protocols that were developed to isotopically label the protein, both uniformly and

site specifically. Chapter 4 describes the heteronuclear multidimensional experiments that were used to facilitate the near complete backbone assignments for this protein, while the final chapter summarizes the results and discusses prospects for future studies.

1.1 NMR Spectroscopy of Proteins

Over the past 15 years, high-resolution NMR spectroscopy has become one of the most powerful techniques for biomolecular structure determination and analysis. The development of pulsed Fourier Transform NMR spectroscopy (Ernst and Anderson, 1966) in conjunction with multidimensional NMR spectroscopy (Jeener, 1971) has made it possible to determine the 3D structure of macromolecules such as proteins and DNA. Using protocols that were developed principally by Wuthrich and his collaborators in the early 1980's, the first 3D protein structure was determined by Williamson et al (1985). Improvements in NMR spectrometer technology such as increased field strength, improved computer speed and increased probe sensitivity coupled with isotopic labelling and improved pulse sequences (Kay and Gardner, 1997; Kay, 1995) have allowed proteins and protein complexes as large as 64 kilodaltons to be characterized by NMR (Shan et al., 1998). To date, NMR spectroscopy has been used to solve more than 1500 biomolecular structures, many of which are available from the Protein Data Bank (PDB) (Bernstein et al., 1977) and Nucleic Acid Database (NDB) (Berman et al., 1992).

Physical principles of NMR. The first commercial NMR spectrometers were available in 1953. The early instruments used permanent magnets rated at 60, 80, 90 or 100 MHz and operated in a continuous wave (CW) or "frequency scan" mode. Historically, the low sensitivity and low resolution of this technique limited the NMR spectroscopy of macromolecules. However, Fourier Transform (FT) NMR (Ernst and Anderson, 1966), introduced in the late 1960s, was shown to be

capable of greatly enhancing the sensitivity of NMR. This sensitivity enhancement was achieved by exciting and subsequently collecting signals from all nuclei (^1H or ^{13}C) in a given molecule simultaneously, rather than sequentially as in the CW mode. An equally important development in the late 1960s was the introduction of superconducting magnets that operated with fields of 200 to 500 MHz. These superconducting magnets improved both sensitivity and resolution of NMR spectrometers. Today, the desire for enhanced sensitivity and resolution still remains and there are now 500, 600, 800, and 900 MHz instruments with helium-cooled superconducting magnets that operate in pulsed FT mode.

Simply stated, an NMR spectrometer (Figure 1.1) consists of: (1) a strong magnet with a very homogeneous field, (2) a radio-frequency (RF) oscillator that generates the electromagnetic (EM) radiation, (3) a RF receiver to pick up changes in EM absorbance or emission, (4) an amplifier to enhance the weak signal emanating from the sample, (5) a recording device and/or computer to convert or deconvolute the data, (6) a sample holder surrounded by a transmitter coil and receiver coil, and (7) a sample spinner that spins the sample to increase the apparent homogeneity of the magnetic field.

NMR spectroscopy involves the measurement of the interaction of magnetic moments of atomic nuclei with the surrounding magnetic field (Kessler et al., 1988). The magnetic moment of a nucleus is associated with the nuclear spin quantum number I . Quantum mechanics states that each nuclear particle (proton or neutron) has a spin value of $I = \frac{1}{2}$. Combinations of multiple particles in the nucleus results in an overall spin value that is specific for each atomic isotope. Isotopes possessing even numbers of protons and neutrons will have $I = 0$ (i.e. ^4He , ^{12}C , ^{16}O). Isotopes having an odd number of protons and an even number of neutrons or vice versa will have $I = \frac{1}{2}$ or a multiple of $\frac{1}{2}$ (i.e. ^1H , ^{13}C , ^{15}N , or ^{19}F). Those isotopes having odd numbers of both protons and neutrons will have $I = 1$ (i.e. ^2H). For NMR spectroscopy of biomolecules, the most important nuclei are those having $I = \frac{1}{2}$ and $I = 1$.

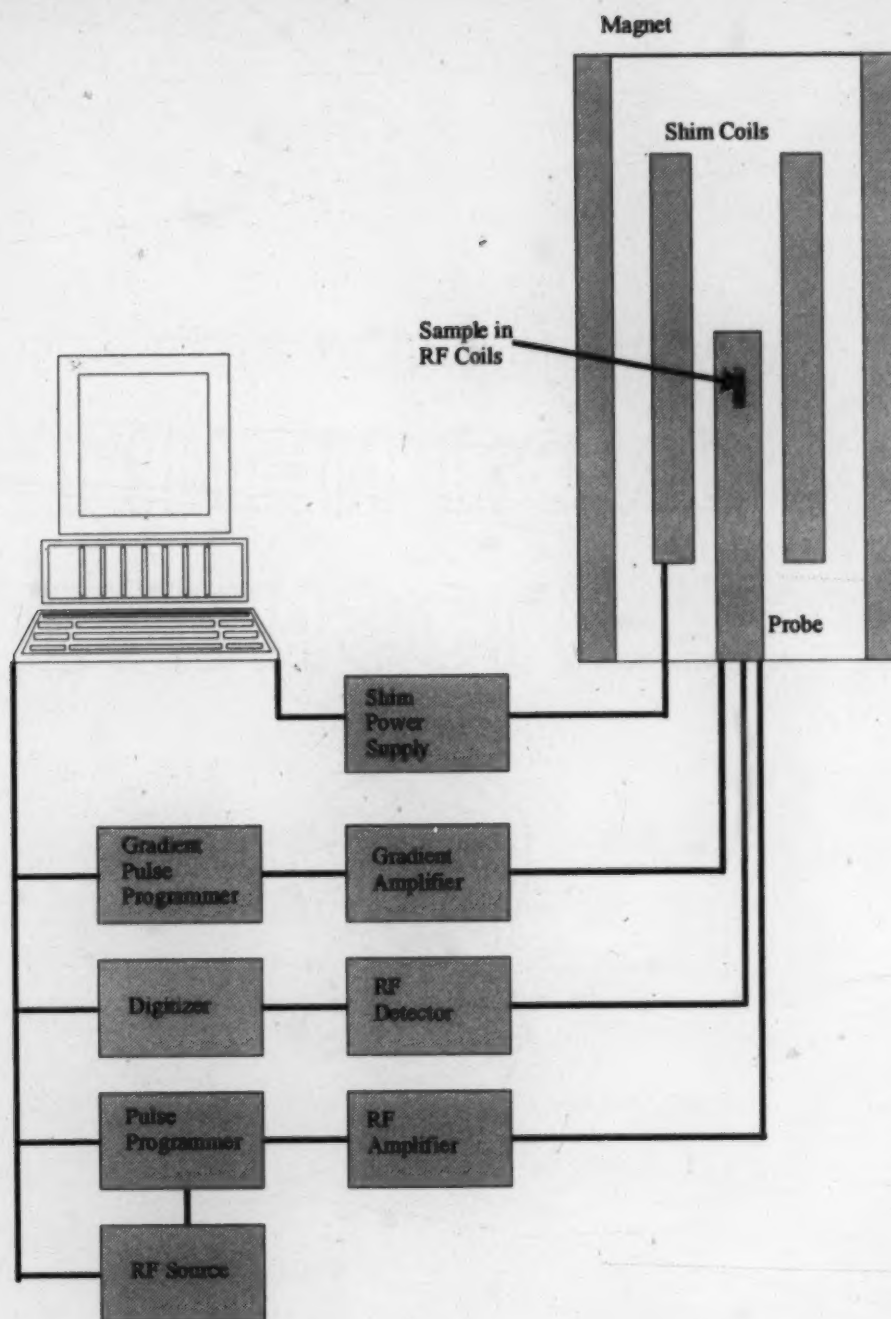


Figure 1.1. Basic outline of a modern, high field NMR spectrometer.

In the presence of an external magnetic field, the spin angular momentum of a nucleus with an overall non-zero spin will undergo a cone-shaped rotational motion called precession (Figure 1.2). The rate of precession for each isotope is referred to as the Larmor frequency (ω_0), which is dependent on the strength of the external field (B_0) and the intrinsic properties of the nucleus reflected in its gyromagnetic ratio (γ). The Larmor frequency may be represented by the following equation:

$$\omega_0 = \gamma B_0 \quad [1.1]$$

Each magnetic nucleus has $2I+1$ possible orientations and $2I+1$ corresponding energy levels with respect to the external magnetic field. Therefore, a nucleus with spin $I = 1/2$ will have two possible orientations which, in the absence of an external magnetic field, will be of equal energy. However, if a magnetic field is applied, the energy levels split (Figure 1.2). The energy difference (ΔE) between these two orientations is directly proportional to the strength of the magnetic field.

$$\Delta E = \gamma h B_0 / 2\pi \quad [1.2]$$

Where h is Planck's constant.

At any given magnetic field, a nucleus can jump from one energy level to the other by absorbing or emitting a discrete amount of energy at its Larmor frequency. In NMR, it is important to remember that one is not looking at a single nucleus, but a population of nuclei ($\sim 10^{20}$). Therefore, at room temperature, the two energy states are unequally populated and the population ratio can be expressed by the Boltzmann equation:

$$N_\beta / N_\alpha = \exp(-\Delta E / kT) \quad [1.3]$$

Where N_α and N_β (Figure 1.2) are the populations of the lower and upper states respectively, k is Boltzmann's constant and T is temperature in degrees Kelvin.

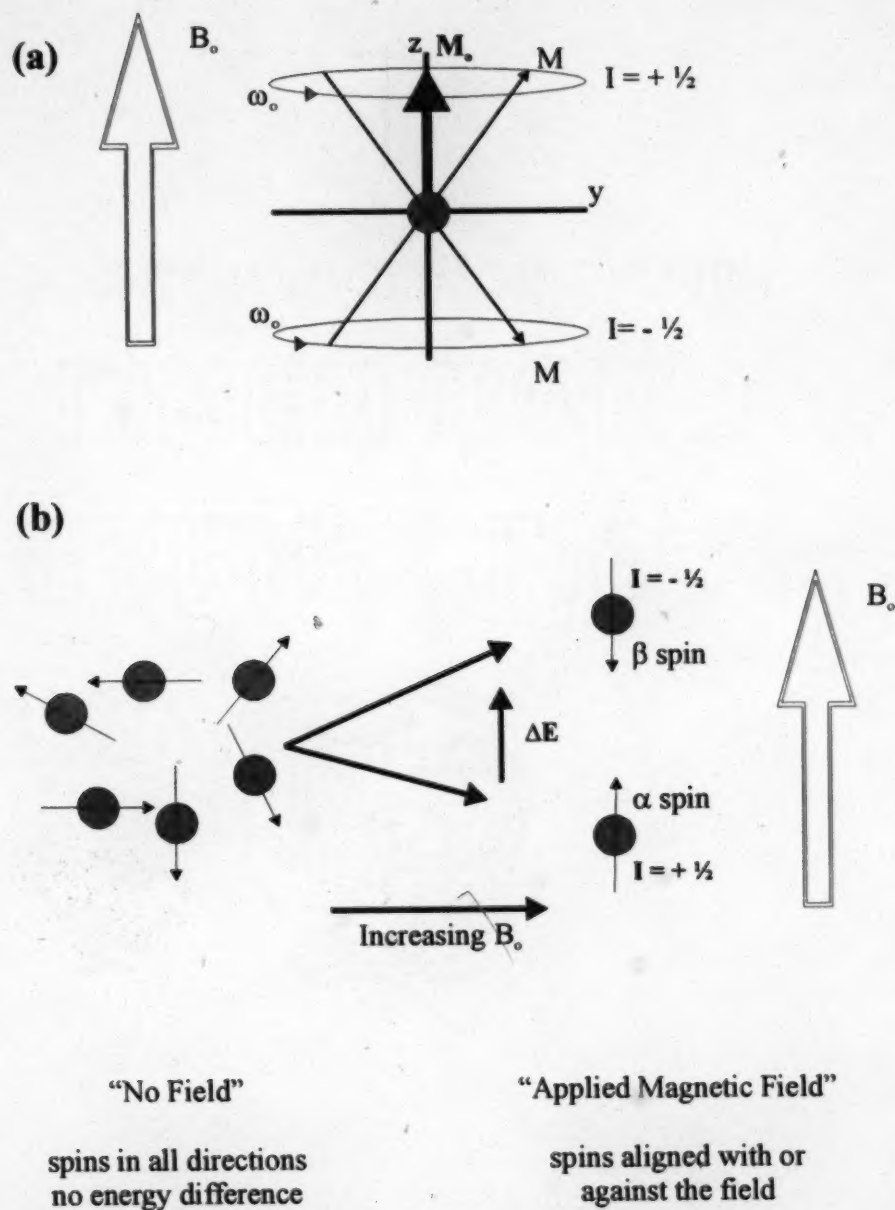


Figure 1.2. Nucleus with spin $I = \frac{1}{2}$ in an applied magnetic field. (a) The two allowed spin states and (b) energy level diagram displaying α and β spin populations.

This population difference generates a net magnetization (M), aligned with the external magnetic field, which will remain static unless the system is disturbed in some way. If a magnetic pulse is applied for a short period of time in such a way that it produces a second magnetic field B_1 perpendicular to the static field B_0 , this pulse will drive M away from its equilibrium position by a so-called "flip angle". This flip angle depends on the time, period and the field strength of B_1 . A 90° pulse or a 90° flip angle is defined as the time it takes a particular field strength to rotate the equilibrium magnetization 90° with respect to its equilibrium direction (the direction of the applied magnetic field, is usually set along the z axis). Having a metal coil in the xy -plane allows the recording of the oscillating current generated by the precessing magnetization. The precessing magnetization eventually returns to equilibrium, with the xy magnetization slowly fading and the z magnetization growing. This oscillating magnetic field can be detected by a coil, converted to an electrical signal and recorded. This signal is called the free-induction decay (FID). Using a mathematical operation called a Fourier transform;

$$f(\omega) = \int f(t) e^{i\omega t} dt \quad [1.4]$$

the FID can then be converted from an exponentially decaying oscillation, which is in the time (t) domain, into the more familiar spectrum, which is in the frequency (ω) domain. This whole process is illustrated in Figure 1.3.

Chemical Shifts. The magnetic field at a nucleus will tend to differ from the applied magnetic field due to the "nuclear shielding" effect of the surrounding electrons.

$$B = B_0 (1 - \sigma) \quad [1.5]$$

Where σ is the shielding constant, B_0 is the applied field and B is the net magnetization.

As electron density increases around a nucleus, so does the "shielding" (σ) and consequently, a stronger magnetic field is required to bring the nucleus into resonance. The shielding constant (σ) is related to the electron density (ρ) at a distance (r) from the nucleus as described by Lamb's formula;

$$\sigma = \frac{4\pi e^2}{3mc^2} \int r\rho(r) dr \quad [1.6]$$

Where m is the mass of an electron (9.11×10^{-31} kg), e is the electron charge (-1.6×10^{-19} Coulombs) and c is the speed of light (2.998×10^8 m/s).

The density of the electrons surrounding a nucleus is a function of the type and proximity of chemical groups neighboring the nucleus of interest. In peptides and proteins, the chemical shifts of nuclei depend on the covalent chemistry of individual amino acids as well as the non-bonded environment (McDonald and Phillips, 1967; Allerhand et al., 1973; de Dios et al., 1993). Involvement in hydrogen bonding and proximity to aromatic and carbonyl groups tend to cause the largest deviations of chemical shifts from the values they would otherwise display in an unstructured (i.e. random coil) peptide (Wishart et al., 1995a).

In practical terms, chemical shifts (δ) are measured by the difference between the frequency (Hz) at which a particular nucleus absorbs (ν_{OBS}) relative to the frequency of a given chemical shift standard (ν_{REF}) such as DSS (2,2'-dimethyl-2-silapentane-5-sulfonate). Chemical shifts are most commonly reported in parts per million (ppm) and are calculated using the equation below.

$$\delta \text{ (ppm)} = \{(\nu_{OBS} - \nu_{REF}) / \nu_{REF}\} \times 10^6 \quad [1.7]$$

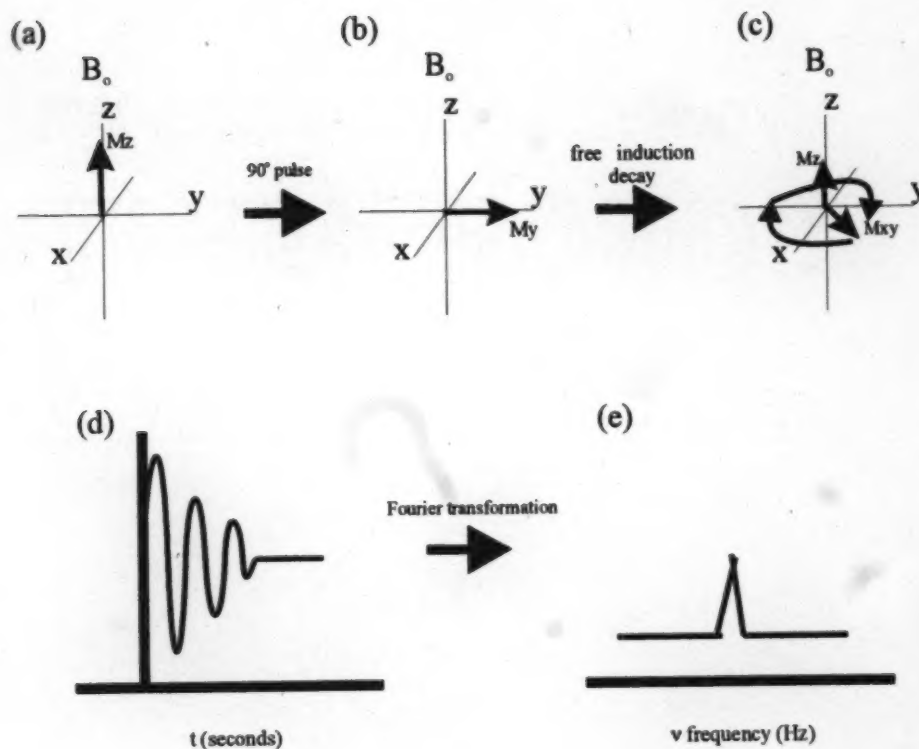


Figure 1.3. Vector picture of pulse NMR. (a) Bulk magnetization (M_z) at equilibrium in magnetic field (B_0), (b) M_z rotates M_x after 90° pulse, (c) M precesses at the Larmor frequency and returns back to equilibrium, (d) free induction decay (FID), (e) fourier transformation of FID yields familiar NMR spectrum.

Chemical shifts are exquisitely sensitive markers of protein structure and protein-ligand interactions. ^1H , ^{13}C and ^{15}N chemical shifts are the most widely studied and the dependence of these shifts on protein secondary structure is well established (Wishart et al., 1991; Wishart and Sykes, 1994b). SAR by NMR (Shuker et al., 1996), a method of screening protein-ligand interactions, also relies on chemical shift information. SAR by NMR relies on measuring the effects of ligands binding to protein targets by analyzing the two-dimensional (2D) ^{15}N - ^1H Heteronuclear Single Quantum Coherence (HSQC) spectrum of the target protein. From perturbations in amide chemical shifts, one can readily identify whether binding has occurred. If the peaks in the 2D ^{15}N - ^1H HSQC have previously been assigned, then the perturbed peaks also indicate the site(s) on the protein where the ligand is binding.

Spin-Spin Coupling. Nuclei that are coupled through chemical bonds exert an influence on each other's effective magnetic field. This effect shows up in a NMR spectrum as a splitting of the main resonance peak and is termed spin-spin coupling, scalar coupling or J coupling. The strength of this interaction is measured by the scalar coupling constant, $^nJ_{ab}$, where n designates the number of covalent bonds separating the two nuclei, a and b. The magnitude of $^nJ_{ab}$ is expressed in Hz and the most important interactions occur when $n = 1$ to 4.

Spin-spin coupling results from the pairing of an observed nucleus to the electrons in its attached bonds. This pairing sensitizes a nucleus to the spin orientation of neighboring nuclei. When the spin of a bonding electron is influenced, it will in turn influence the spin of other bonding electrons that will then affect an adjacent nucleus and so on down the length of the molecule. This coupling cascade is the basis of through-bond NMR experiments such as COSY and TOCSY.

For proteins, homonuclear experiments such as TOCSY and COSY typically depend on $^3J_{\text{HH}}$ -couplings (such as backbone ^{15}N to backbone H_α) to

effect magnetization (polarization) transfer between spins. Significant magnetization transfer via these couplings requires a delay on the order of $1/J$. Thus, smaller J coupling values ($^3J_{HH} = 3-10$ Hz versus $^1J_{CH} = 125-160$ Hz) require longer periods of time to transfer magnetization ($^3J_{HH} = 0.33-0.10$ s versus $^1J_{CH} = 0.008-0.006$ s). This can be problematic in the NMR spectroscopy of larger proteins that possess more efficient relaxation for their NMR active nuclei. An advantage of heteronuclear NMR spectroscopy is the efficient magnetization or polarization transfer enabled by using one- and two-bond J_{CH} , J_{NH} , J_{CC} , and J_{CN} couplings. This concept will be explored in more detail in the next section, which discusses triple resonance spectroscopy.

It is important to note that structural information can also be derived from these coupling constants. As was first described by Karplus (1959), the magnitude of a $^3J_{ab}$ coupling constant contains information about the dihedral angle (θ) formed by the three covalent bonds between nuclei a and b . This relationship is characterized by the following equation:

$$^3J_{ab} = A \cos^2 \theta + B \cos(\theta) + C \quad [1.8]$$

Where A , B , and C are empirically derived constants for each type of coupling constant. $^3J_{H^aNH^b}$ and $^3J_{H^aCH^b}$ are the two main coupling constants that provide structural information about dihedral angles in peptides and proteins (Pardi et al., 1984; Bystrov, 1976; Wang and Bax, 1995 and 1996).

Spin Relaxation and Linewidth. Spin relaxation is defined as the return to equilibrium of the projection of the magnetic moment of a large number of spins after an initial perturbation or excitation. This is demonstrated in Figure 1.3(c), which shows transverse magnetization (M_y) returning to its equilibrium magnetization (M_0) along the magnetic field axis (longitudinal - M_z) and perpendicular to it (transverse - M_{xy}). This phenomenon is of great importance to

NMR spectroscopy as the measure of relaxation rates can provide information about the motion or dynamics of the molecules being studied (Kay et al., 1989).

For spin $I = \frac{1}{2}$ nuclei in proteins (such as ^1H , ^{13}C and ^{15}N), the two prominent relaxation mechanisms are chemical shift anisotropy (CSA) and dipole-dipole (DD) interactions. As stated earlier, chemical shifts reflect the local magnetic fields experienced by different nuclei. These local fields are anisotropic and consequently the components of the local fields vary as the molecule reorients itself as a result of molecular motion. These varying magnetic fields are what cause CSA relaxation. The second important relaxation mechanism arises from dipole-dipole (DD) interactions. Any NMR active nucleus in a molecule generates a magnetic dipolar field that is proportional to the magnetic moment of the nucleus. As a molecule tumbles in solution, this field fluctuates constituting a mechanism for relaxation of nearby spins.

Dipolar relaxation plays a prominent role in the NMR spectroscopy of proteins. ^1H relaxation is dominated by dipolar interaction between nearby protons that are within approximately 5 Å. Similarly, the relaxation of protonated ^{13}C and ^{15}N nuclei also arises primarily from dipolar interactions with directly bonded protons. CSA plays a secondary role in the relaxation of most nuclei. However, it is an important relaxation mechanism for nuclei with large chemical shift ranges. Therefore ^{13}C , ^{15}N and ^{31}P nuclei experience significant CSA contributions to relaxation whereas protons experience negligible effects.

Both CSA and DD interactions play a role in the macroscopically observable relaxation rates known as spin-lattice (longitudinal or T_1) and spin-spin (transverse or T_2) relaxation. T_1 relaxation is the measure of the efficiency with which excited nuclear spins (spins that have experienced a 90° pulse) return to thermal equilibrium by exchanging energy with the "lattice" or surroundings. T_2 relaxation describes the loss of transverse magnetization due to the exchange of spin energies between excited spins. In general, T_1 is always greater than or equal

to T_2 because there are additional contributions to transverse relaxation such as non-uniformity of the static magnetic field, B_0 . When the sample is divided into regions such that the field is uniform (regions known as isochromats), then the total magnetization is the sum of these regions. Inherently, there will be differences in sample uniformity that blur the bulk magnetization leading to inhomogeneous line broadening or enhanced transverse relaxation.

In NMR, linewidth ($\Delta\nu_{1/2}$) is defined as the full-width at half-height of the resonance line-shape (Figure 1.4, p.17). Linewidth is a key factor affecting both resolution and signal-to-noise ratio of NMR spectra. Spectral linewidths are intricately related to T_2 relaxation, which is approximately proportional to the overall rotational correlation time (τ_c) of the molecule (Figure 1.5, p.18). τ_c is defined as the approximate average time for the molecule to rotate by one radian (57°). Note that τ_c varies with molecular size, molecular shape, solvent viscosity and temperature (Cavanagh, 1996).

Table 1.2 outlines the relationships between, molecular weight (MW), rotational correlation time (τ_c), backbone ^{15}N - and $^{13}\text{C}_\alpha$ - T_2 relaxation times and half-height linewidths ($\Delta\nu_{1/2}$). These values are estimated by the following convenient equations (Spyracopoulos et al., 1999). The rotational correlation time (τ_c) of a spherical protein (in nanoseconds) can be estimated from the MW (in kilodaltons) of the protein via

$$\tau_c = \frac{1}{2} \times \text{MW} \quad [1.9]$$

In addition, the transverse relaxation times (in seconds) of backbone ^{15}N and $^{13}\text{C}_\alpha$ can be conveniently estimated from the correlation time (in nanoseconds) via the following equations.

$$1/T_{2,\text{NH}} = 1.11\tau_c \quad [1.10]$$

$$1/T_{2,CH} = 4.32 \tau_c \quad [1.11]$$

Estimates of the half height linewidths ($\Delta\nu_{1/2}$) are calculated from the T_2 values using the following equation.

$$\Delta\nu_{1/2} = 1/\pi T_2 \quad [1.12]$$

As can be noted in Table 1.2, increased molecular weight corresponds to increased transverse relaxation rates (more rapid decay of transverse magnetization) and broader linewidths. Consequently, larger proteins experience greater sensitivity losses (signal loss) during NMR data acquisitions when compared to smaller proteins. This loss of sensitivity accounts, in large part, for the molecular weight limit imposed on solution NMR methods.

Table 1.2: Relationships between molecular weight, correlation time (τ_c), backbone ^{15}N - T_2 , $^{13}\text{C}_\alpha$ - T_2 , and half-height linewidths ($\Delta\nu_{1/2}$) of backbone ^{15}N and $^{13}\text{C}_\alpha$ nuclei.

MW (kDa)	τ_c (ns)	^{15}N - T_2 (s)	$^{13}\text{C}_\alpha$ - T_2 (s)	^{15}N - $\Delta\nu_{1/2}$ (Hz)	$^{13}\text{C}_\alpha$ - $\Delta\nu_{1/2}$ (Hz)
5	2.5	0.360	0.093	0.88	3.42
10	5	0.180	0.046	1.76	6.91
24	12	0.075	0.019	4.24	16.75
48	24	0.037	0.010	8.60	31.83

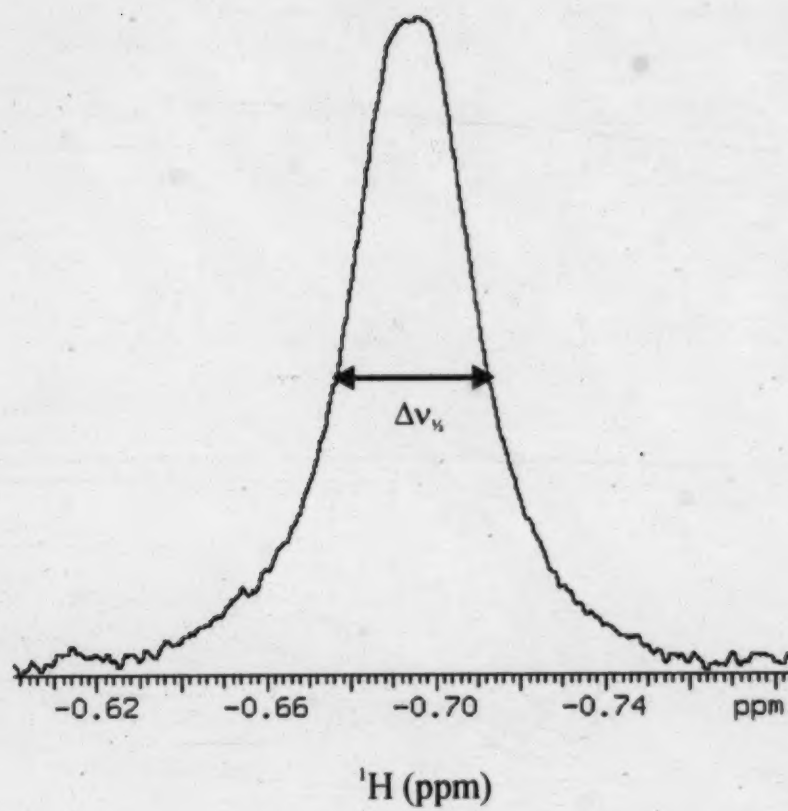


Figure 1.4. A resonance in a 1D ^1H NMR spectrum displaying half-height linewidth ($\Delta\nu_{1/2}$).

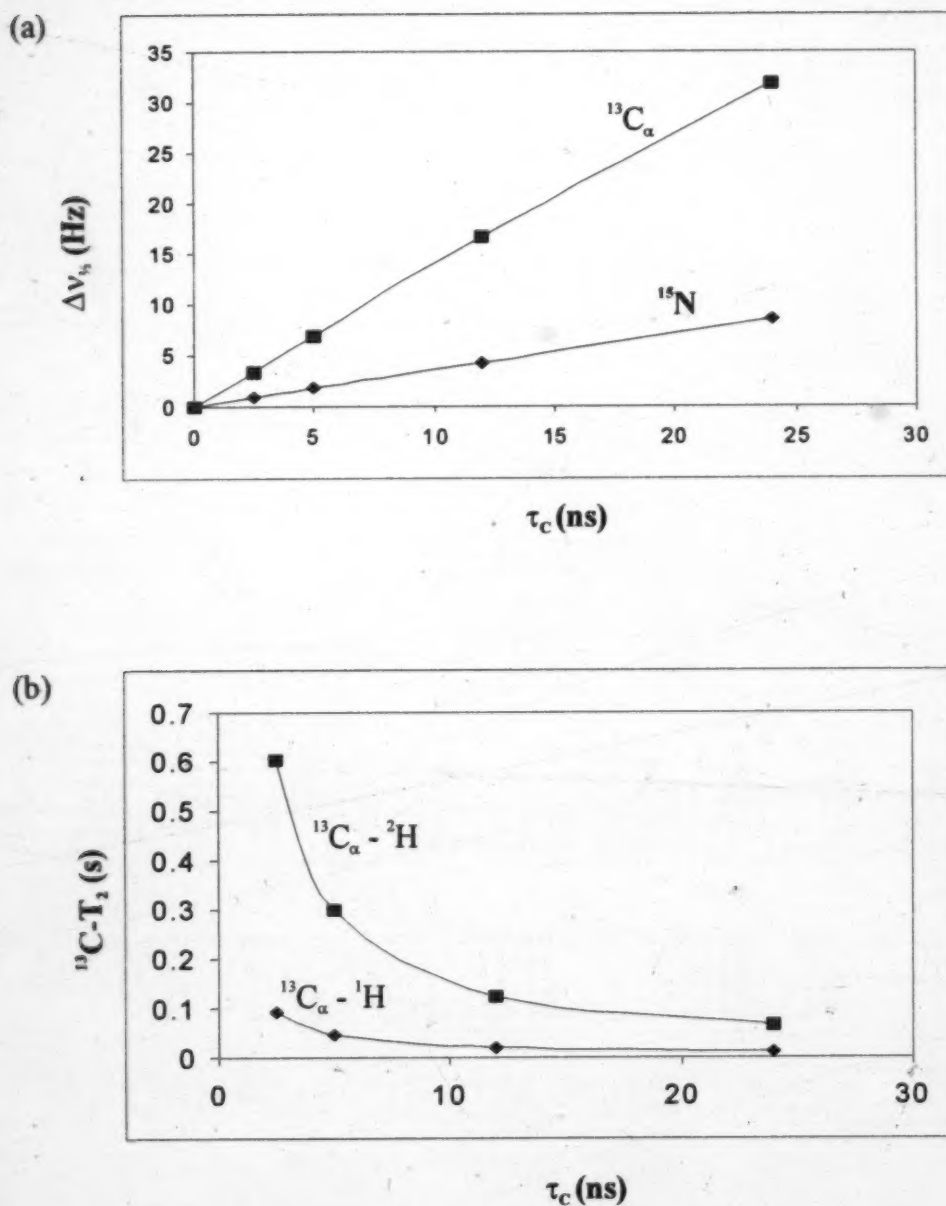


Figure 1.5. Relationships between τ_c , T_2 and $\Delta\nu_N$. (a) Backbone $^{13}\text{C}_\alpha$ - and ^{15}N - $\Delta\nu_N$ in relation to τ_c , (b) backbone $^{13}\text{C}_\alpha$ - T_2 when attached to ^1H and ^2H in relation to τ_c .

1.2 Sequence-Specific Assignment of Larger Proteins

In NMR, the first step in determining the three dimensional structure of a protein is to assign as many resonances as possible to specific nuclei in the individual amino acid residues. For proteins of approximately 100 amino acids or less, two-dimensional ^1H NMR experiments developed during the 1970s and 1980s are usually sufficient to permit unambiguous assignment (Wuthrich, 1986). However, due to problems of overlap and sensitivity, ^1H homonuclear NMR spectroscopy is usually insufficient for proteins having a molecular weight exceeding 10 kilodaltons. To permit the sequential assignments of larger proteins, a spectroscopist must often resort to isotopic labelling and very careful sample preparation to overcome three major problems: 1) protein aggregation, 2) spectral overlap and 3) loss of sensitivity. The following three sections will briefly discuss strategies and techniques that have evolved over the past 10 years in order to overcome these three obstacles.

Protein Conditioning. A critical step in the sequential assignment process is the acquisition of "good" spectra to facilitate analysis. Many proteins tend to aggregate far below the concentrations required for NMR experiments. This aggregation is problematic in that it leads to significantly broadened lines, poor resolution and loss of signal, all of which complicate spectral analysis. Possible reasons for aggregation include electrostatic or hydrophobic interactions present between protein molecules (Wagner, 1993). "Protein conditioning" is a term that encompasses the strategies used to prevent protein aggregation in NMR samples (Wagner, 1993). Typically, these strategies include the screening samples where the pH, temperature, buffer, and ionic strength are varied in order to determine the appropriate conditions where the protein behaves as a monomer. There have also been instances where the addition of detergents, such as CHAPS (Anglistter et al., 1993), and the introduction of point mutations (Kaarsholm and Ludvigsen, 1994) have been shown to prevent protein aggregation.

Isotopic Labelling. Because of the limited chemical shift dispersion of protons (~ 10 ppm), the complexity of 2D ^1H homonuclear spectra increase roughly as the square of the molecular weight. If it was possible to either increase the chemical shift dispersion or the "dimensionality" of the spectra, this problem of spectral complexity or spectral overlap could be greatly reduced. For this reason isotopic labelling with ^{13}C , ^{15}N (both have large chemical shift ranges), and in some instances ^2H , has become standard practice for larger proteins. Normally, these isotopes must be incorporated at levels greater than 90% to achieve satisfactory results (Mossakowska and Smith, 1997). Furthermore, by uniformly or site specifically labelling a protein with either ^{13}C or ^{15}N , a set of nuclei other than ^1H can be detected by NMR. This adds new dimensionality to the experiments, thereby reducing spectral overlap and facilitating spectral analysis (Figure 1.6, p.21).

In order to achieve this high degree of isotopic labelling, one must have a high-yield expression system (usually bacterial) in which the cells can grow efficiently on minimal nutrient media composed of simple organic carbon sources (such as ^{13}C enriched glucose), nitrogen sources (such as ^{15}N enriched ammonium chloride) and ^2H in the form of $^2\text{H}_2\text{O}$ (Reid et al., 1997). Such a biosynthetic system can be used to either uniformly label proteins or to site-specifically label them with individual amino acids (McIntosh et al., 1990; LeMaster and Richards, 1985).

Because of the beneficial effects of line narrowing, deuterium labelling has played a beneficial role in the solution NMR studies of larger macromolecules (Gardner and Kay, 1998). For heteronuclei (such as ^{13}C and ^{15}N) that are directly bonded to protons, the major source of relaxation is due to dipolar fields caused by the ^1H spins (Kay, 1997). The size of these dipolar fields is related to the gyromagnetic constants of the nuclei responsible for them. By substituting ^{13}C -bound protons for deuterons, the size of these dipolar fields can be reduced by a

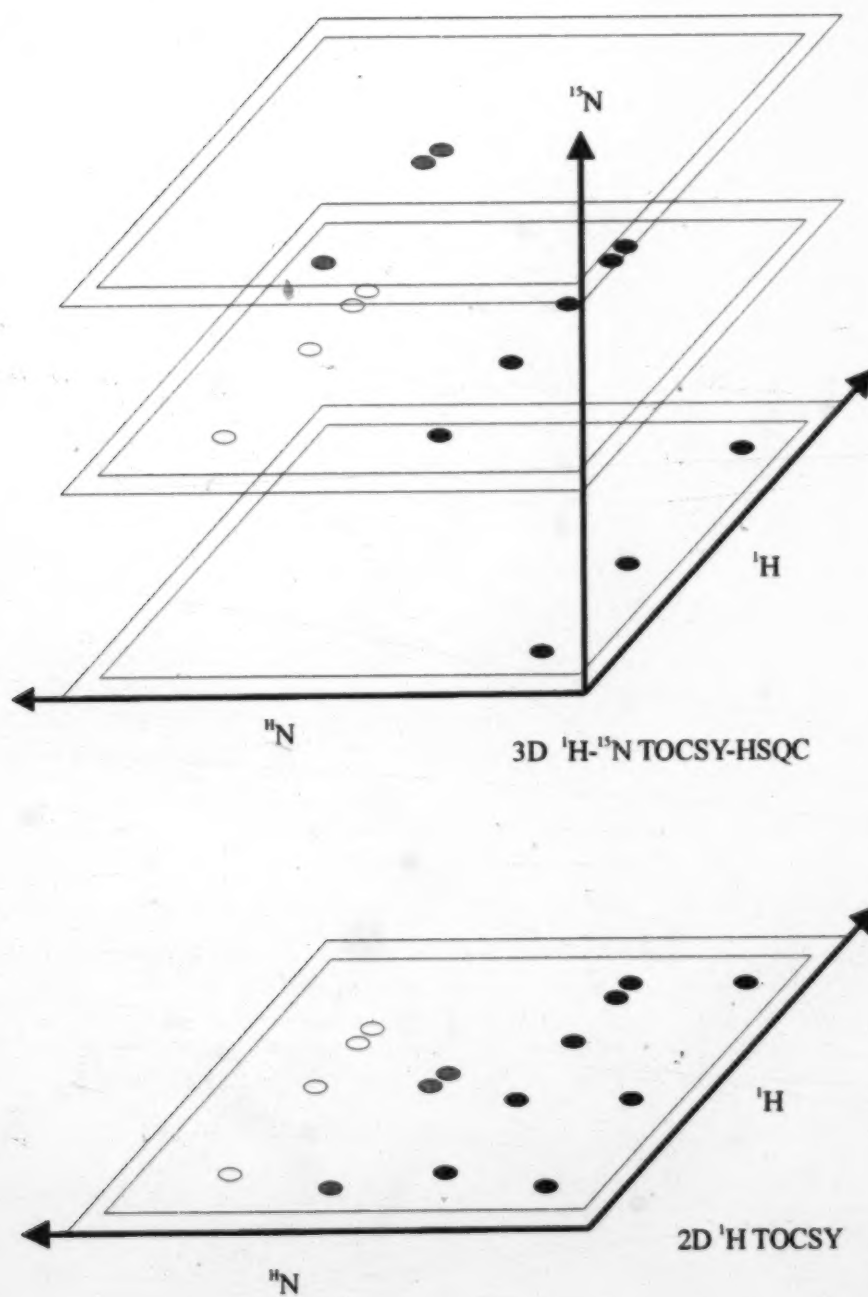


Figure 1.6. Adding a third dimension to NMR experiments.

factor of ~ 6.5 . This is because the gyromagnetic constant of deuterium is ~ 6.5 times smaller than that of hydrogen. This results in a substantial decrease in the ^{13}C relaxation rate. In addition, any remaining protons in the protein will also experience a decrease in their T_1 and T_2 decay rate. This is because some of the neighboring protons causing dipolar relaxation (those within $5 \text{ \AA}$) have been replaced by deuterons. Figure 1.5 (p.18) illustrates the increase in backbone $^{13}\text{C}_\alpha$ carbon T_2 's that can be expected upon deuteration (Kay, 1997). These larger T_2 values correspond to slower relaxation rates and therefore a concomitant narrowing in linewidths for the affected nuclei.

Triple Resonance NMR Spectroscopy. Many of the NMR experiments designed for the analysis of larger proteins require uniform labelling with ^{15}N and ^{13}C . These experiments are called triple resonance techniques because three resonances (^1H , ^{13}C and ^{15}N) are recorded in a single experiment. Triple resonance techniques represent a significant advance over homonuclear ^1H techniques because they exploit the large couplings that exist between the ^{15}N and ^{13}C nuclei and their directly bonded protons. They also exploit the sensitivity enhancement that can be achieved with polarization transfer techniques such as INEPT (Morris and Freeman, 1979) which enhance the signals that can be detected from low sensitivity nuclei such as ^{15}N and ^{13}C .

A typical multidimensional NMR experiment can be represented as a series of magnetization or polarization transfer steps, such as:

$$A \rightarrow B \rightarrow C \rightarrow D \rightarrow \dots \rightarrow Z$$

where the transfer of magnetization (through combinations of INEPT and reverse INEPT pulses) proceeds from A to B to C and so on until it reaches Z where signal is acquired. The amount of time necessary to transfer significant magnetization along each link in the chain is a function of the strength of the coupling between the two nuclei. This time is given by $1/J$ (s). Therefore larger couplings lead to

more efficient transfer of magnetization. During this transfer, the signal decays via relaxation processes that become more rapid as the size of the protein increases. If the decay rate is on the same order as the magnetization transfer rate, sensitivity of the experiments will be seriously compromised.

Advantages of Triple Resonance NMR Spectroscopy. For ^1H homonuclear NMR experiments, $^3J_{\text{HH}}$ coupling (approximately 5 – 12 Hz) are normally used to effect magnetization transfer between spins in the protein. In these cases, significant magnetization transfer via this coupling would require 80 – 200 ms delays. During this time, the efficient relaxation rates of nuclei in large proteins would lead to significant loss of signal intensity.

In contrast to ^1H homonuclear NMR, triple resonance NMR spectroscopy utilizes two techniques to increase signal at the end of magnetization transfer. First, triple resonance NMR utilizes one and two bond $^1J_{\text{CH}}$, $^1J_{\text{NH}}$, $^1J_{\text{CC}}$, $^1J_{\text{CN}}$ and $^2J_{\text{CN}}$ couplings to effect efficient magnetization transfer. Using these couplings, the approximate time necessary for efficient magnetization transfer between backbone nuclei can easily be approximated as $1/J$ (Table 1.3) (Delaglio et al., 1991; Powers et al., 1992). Therefore, efficient magnetization transfer is the first mechanism by which triple resonance NMR experiments optimize signal intensity.

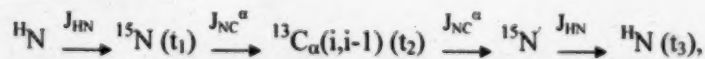
Table 1.3. Coupling constants and coherence transfer rates for ^1H - ^1H homonuclear, ^{15}N -edited and triple resonance NMR experiments

J coupling	Coupling Constant (Hz)	Coherence Transfer Time (s)
$^3J_{\text{H}\alpha\text{H}\beta}$	~ 3 – 11	0.333 – 0.091
$^1J_{\text{NH}}$	~ 95	0.011
$^2J_{\text{NH}\alpha}$	~ 3 – 11	0.333 – 0.091
$^1J_{\text{NC}\alpha}$	~ 9 – 13	0.111 – 0.076
$^2J_{\text{NC}\alpha}$	~ 5 – 10	0.200 – 0.100
$^1J_{\text{NCO}}$	~ 15	0.067
$^1J_{\text{CaCO}}$	~ 55	0.018
$^1J_{\text{CC}}$	~ 30 – 40	0.033 – 0.025
$^1J_{\text{CH}}$	~ 140	0.007

The second mechanism to optimize signal intensity involves removing the primary source of relaxation in proteins, the protons. Labelling a large protein with deuterium plays an adjunctive role to triple resonance experiments to increase sensitivity of NMR experiments. In the context of pulse sequences, the rate at which a resonance signal decays is of great significance to the success of any NMR experiment. For triple resonance experiments, deuterium (bonded to ^{13}C nuclei) can significantly reduce the relaxation rate of ^{13}C nuclei thereby decreasing the rate at which the ^{13}C signal decays. This reduction in relaxation rate, in combination with efficient magnetization transfer via large one and two bond couplings, ensures that signal is still discernible at time of acquisition. Indeed, many triple resonance pulse schemes have been "deuterium modified" to take advantage of the reduced ^{13}C relaxation rates in highly deuterated proteins (Gardner and Kay, 1998; Yamazaki et al., 1994a/b).

A prototypical triple resonance NMR experiment. A useful triple resonance experiment for sequential assignments of larger proteins is called the HNCA experiment (Kay et al., 1990c; Grzesiek and Bax, 1992). A description of this experiment serves to illustrate the relevant features of heteronuclear multidimensional experiments. As the name implies, this experiment provides protein backbone information by correlating amide proton ($^1\text{H}_\text{N}$), attached amide ^{15}N and intraresidue and preceding residue $^{13}\text{C}_\alpha(i, i-1)$ chemical shifts.

The flow of magnetization transfer in this experiment can be described concisely by



where the relevant couplings involved in each of the magnetization transfer steps are indicated above the arrows and t_1 , t_2 and t_3 denote acquisition times.

This can be more formally denoted using product operator notation as follows:

$$I_z J_z K_z \rightarrow I_y J_z K_z \rightarrow I_z J_y K_z [t_1] \rightarrow I_z J_z K_y [t_2] \rightarrow I_z J_y K_z \rightarrow I_y J_z K_z [t_3].$$

Where I, J, K represent ^1H , ^{15}N , and $^{13}\text{C}_\alpha$ respectively and t_1 , t_2 and t_3 are acquisition periods for ^{15}N , $^{13}\text{C}_\alpha$ and ^1H chemical shifts respectively. The y and z subscripts describe when the nucleus' magnetic moment is transverse and longitudinal to B_0 respectively.

In the HNCA experiment (Figure 1.7, p.26), magnetization is first generated on the amide proton ($^1\text{H}_\text{N}$) ($I_y J_z K_z$). An INEPT pulse sequence transfers magnetization to ^{15}N via $^1J_{\text{NH}}$ (~ 90 Hz) coupling. ^{15}N chemical shifts ($I_z J_y K_z$) are recorded over the acquisition time t_1 . A second INEPT pulse sequence is used to transfer magnetization to the intrasidue and the preceding residue $^{13}\text{C}_\alpha$ via $^1J_{\text{NC}\alpha}$ (8 – 12 Hz) and $^2J_{\text{NC}\alpha}$ (~ 7 Hz) couplings respectively. The two $^{13}\text{C}_\alpha$ chemical shifts ($I_z J_z K_y$) are recorded over the acquisition period t_2 . Magnetization is then transferred back to the protons using two reverse INEPT sequences and the proton chemical shift ($I_y J_z K_z$) is collected over the third and final acquisition period t_3 . Fourier transformation of the resultant FID yields a frequency domain map with both intra- and interresidue correlations between $^{13}\text{C}_\alpha$, ^{15}N , and $^1\text{H}_\text{N}$.

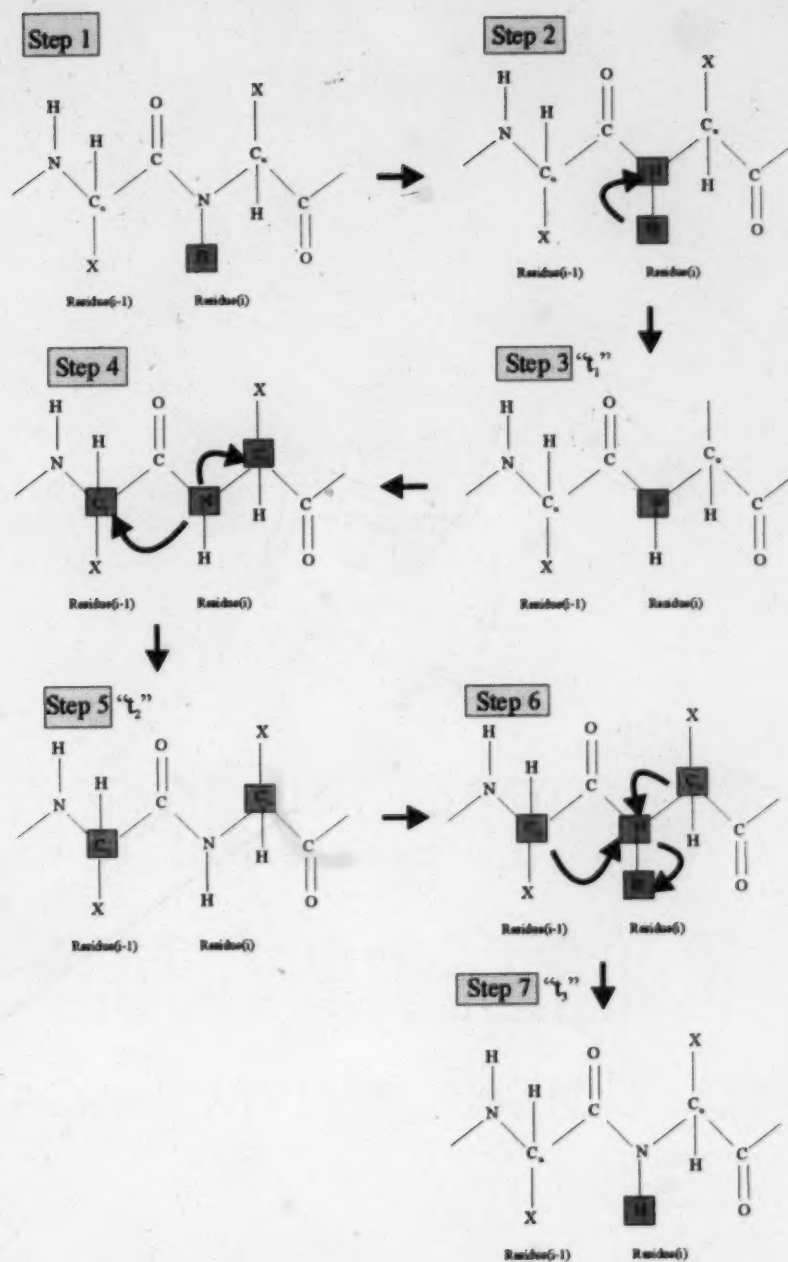


Figure 1.7. Outline of magnetization transfer in HNCA experiments.

1.3 Hepatitis A Virus 3C Protease

Viral Hepatitis A. Hepatitis A is one of the most common forms of acute viral hepatitis in the world. Seroprevalence rates approaching 100% have been noted in the economically developing countries of Africa, Asia, and Latin America. On the other hand, because of improved sanitation, hepatitis A rates have been declining in North America, western Europe and many Mediterranean countries (Koff, 1998). Nonetheless, in 1992 – 1994, of all notifiable infectious diseases in the USA, hepatitis A ranked third amongst children and adolescents, and fifth and sixth amongst men and women respectively (<http://www.cdc.gov/>). Hepatitis A usually begins with the abrupt onset of non-specific, premonitory constitutional and gastrointestinal symptoms. Symptoms include a combination of fever, malaise, weakness, anorexia, nausea, vomiting, arthralgias (joint pain), and myalgias (muscle pain). Unlike hepatitis B or C, hepatitis A is not linked to chronic liver disease and it does not result in persistent viraemia or an intestinal carrier state. Rather, infection is maintained by serial transmission from acutely infected individuals (Koff, 1998). While effective vaccines for Hepatitis A have recently been introduced (Thiel, 1998), there is no available drug therapy at this time. In most cases, Hepatitis A infections are relatively harmless. However, if individuals are co-infected with chronic hepatitis, the hepatitis A infection is often more dangerous (Sjorgren, 1998; Vento et al., 1998).

Picornaviridae. Hepatitis A virus (HAV) is a member of the picornavirus family (Jia et al., 1991). More than 200 known viruses belong to this family including many that are responsible for human disease (Rueckert, 1996). Each member of the picornavirus family shares the same major events of the viral replication cycle, including specific “self” proteolytic processing of the viral polyprotein (Palmenberg, 1990). Differences in viral replication and polyprotein processing have been used to differentiate the Picornaviridae family into six different genera (Table 1.4).

Table 1.4. The Picornaviruses.

Genus	Number of Serotypes	Viruses	Diseases caused in Humans
Enterovirus	> 90	Poliovirus Coxsackievirus Echovirus	Intestinal infections Poliomyelitis Myocarditis Meningitis Encephalitis
Rhinovirus	> 100	Rhinovirus	Common cold
Aphthovirus	7	FMDV Equine rhinovirus	Foot and mouth disease of cloven-hoofed animals
Cardiovirus	2	Encephalomyocardiovirus Theiler's murine encephalitis virus	None known
Hepatovirus	1	Hepatitis A virus	Infectious hepatitis
Parechovirus	2 (perhaps 3)	Parechovirus 1 Parechovirus 2 Ljungan river virus	Myocarditis Intestinal infections

Polyprotein Processing. At the time of the initial infection, the Hepatitis A virus will attach itself to a host cell via specific cell surface receptors and undergo conformational changes that allows it to release its genome into the cytosol of the host cell (Bergmann and James, 1999). Translation of the viral single-stranded RNA yields a large ~ 251 kDa polyprotein (Linemeyer et al., 1985; Najarian et al., 1985) that is co-translationally and sequentially cleaved into individual proteins (Palmenberg, 1990). Once these proteolytic cleavages have been performed, the capsid proteins undergo the conformational changes necessary to allow them to assemble into the procapsid and eventually into the provirion (Rueckert, 1996). A simplified scheme describing the polyprotein processing is shown in Figure 1.8 (p.30).

In hepatoviruses, the 3C gene product is the only protease responsible for polyprotein processing as it is known to cleave all the specific cleavage sites within the polyprotein (Graff et al., 1999; Martin et al., 1999; Jia et al., 1993). The HAV 3C protease typically cleaves between a glutamine residue (P1) and a smaller residue (P2) such as alanine, serine, or glycine (Petithory et al., 1991). Residues in the P4 to P2' positions (following the nomenclature of Schechter and Berger 1967) play a role in cleavage specificity, but a glutamine in the P1 position of the substrate is the most important determinant (Figure 1.9, p.31) (Jewell et al., 1992). X-ray crystallographic studies of the HAV 3C protease (Bergmann et al., 1997) indicate that the other substrate positions of importance are: (1) P2 being occupied by a serine, threonine (2) P3 being occupied by phenylalanine, tryptophan, or glutamate, and (3) P4 being occupied by a large hydrophobic aliphatic residue such as leucine, isoleucine, or valine.

Structure of HAV 3C protease. The structure of the Hepatitis A Virus 3C protease was initially determined by X-ray crystallography in 1994 (Allaire et al., 1994) and later refined by Bergmann et al., in 1997. As can be seen in Figure 1.10 (p.34), HAV 3C protease is composed of two antiparallel β -barrel domains with the active-site and substrate-binding region lying in the surface groove between the two β -barrel domains (Bergmann et al., 1997). Structural comparisons indicate that the 3C proteases belong to the superfamily of chymotrypsin-like serine proteases (Gorbalenya and Snijder, 1996). In this structural family, both domains participate in the binding of peptide substrate. The N-terminal domain participates in binding residues following the scissile peptide bond (P1' to P3') whereas certain C-terminal domains form the pockets that interact with the residues in the P2 to P4 positions of the substrate (Perona and Craik, 1995).

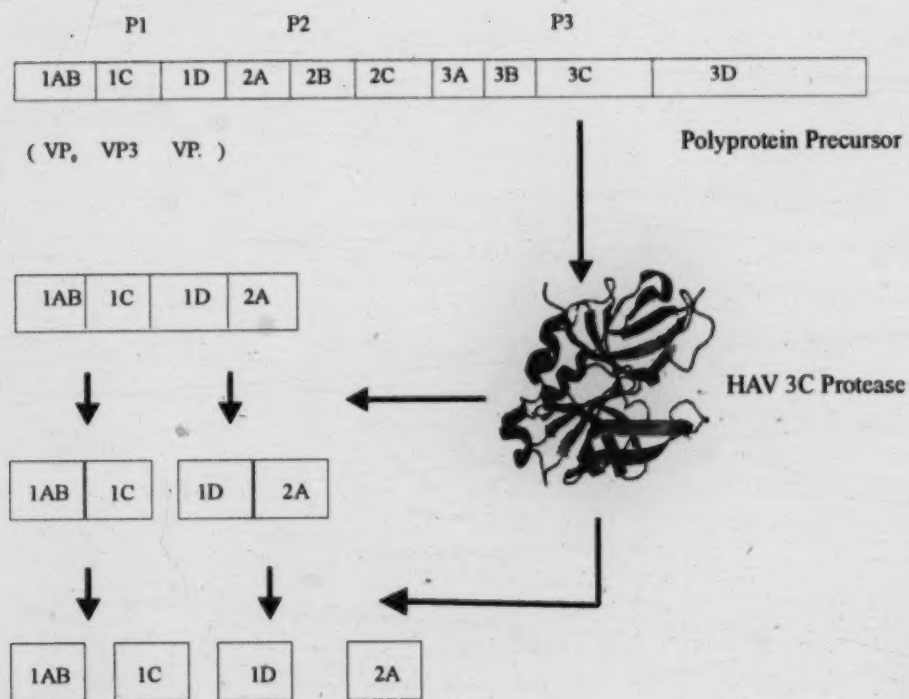


Figure 1.8. Simplified scheme of polyprotein processing by HAV 3C Protease.

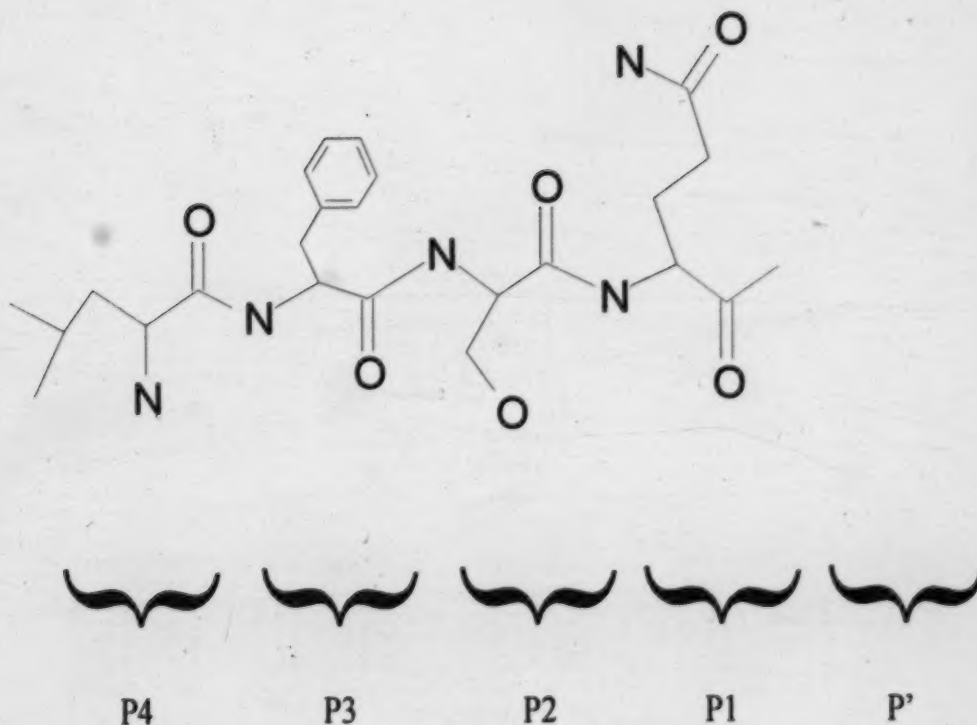


Figure 1.9. Substrate specificity of HAV 3C protease. Cleavage preferences have been shown for P1 - P4 positions although HAV 3C protease is much less specific for P' residues (Bergmann et al., 1997).

Mechanism of HAV 3C protease. In general, there are three chemical groups that function in the active sites of proteolytic enzymes (James, 1993). These include: (1) a nucleophile that attacks the carbonyl of the scissile peptide bond, (2) a general acid-base catalyst that assists in the attack and protonates the leaving group, and (3) an electrophilic structure (oxyanion hole) that stabilizes the negative charge that develops on the carbonyl (Bergmann and James, 1999; James, 1993).

The active site residues of the 3C protease are shown in Figure 1.11 (p.35). Cys172 and His44 form the cysteine-histidine dyad (nucleophile-acid base catalyst) while the oxyanion hole (Pro169, Gly170, Met171 and Cys172) is just above the dyad and to the right (Bergmann et al., 1997). This catalytic-triad arrangement is similar to that observed in the chymotrypsin-like serine proteases but due to the large size of the cysteine sulfur atom, the 3C protease active site is larger (Bergmann and James, 1999). While HAV 3C protease seems to have a mechanism or structure similar to chymotrypsin-like serine proteases, the presence of a cysteine instead of a serine appears to make it something of a hybrid cysteine/serine protease.

HAV 3C protease has four pockets (S1 – S4) that interact with four peptide substrate positions (P1 to P4). The S1-binding pocket is comprised of the main chain atoms of Ala193, Gly194, Gly167, and Leu168 along with the side chains of Leu199 and Met171 and the imidazole of His191. The side chains of Tyr143, His145, and Ala193 form the S2 pocket while the S3 pocket is thought to be formed by Val144, Lys146. The S4 pocket is formed by the side chains of A141, Val144, A198 and Val200 (Bergmann et al., 1997).

The active site residues and the residues comprising the S1 through S4 binding pockets will be very important for future "SAR by NMR" studies of the HAV 3C protease. Using 2D ^{15}N - ^1H HSQC experiments it should be possible to note where inhibitors are binding to the protease. It is hoped that compounds will be found that cause perturbations to the resonances corresponding to the active site

residues as well as to those resonances corresponding to the S1 through S4 binding pockets.

Inhibitors of the Hepatitis A Virus 3C Protease. Inhibitors of the 3C proteases are generally comprised of a chemical group that covalently attaches to the nucleophilic thiol in the active site and a second group that blocks some other interactions between the protease and its substrates (Figure 1.12, p.36). Typical cysteine protease inhibitors are iodoacetamide, N-ethylmaleimide, epoxides, and aldehydes (Malcolm, 1995). However, recently it was shown that the fluoromethylketones and γ -aminovinylsulfones (Rasnik, 1996) display greater promise than typical cysteine protease inhibitors. A combination of a fluoromethylketone and peptide that mimics the natural peptide specificity such as Acetyl-Leu-Ala-Ala-Gln-FMK has been shown to be an effective inhibitor to the HAV 3C protease both in vitro and in vivo (Morris et al., 1997) (Figure 1.13). Other types of compounds that are being studied as potential inhibitors include α,β -unsaturated carboxylesters, β - and γ -lactones, lactams, isatins (2,3-dioxindoles), and triterpene sulfates (Skiles and McNeil, 1990; Brill et al., 1996).

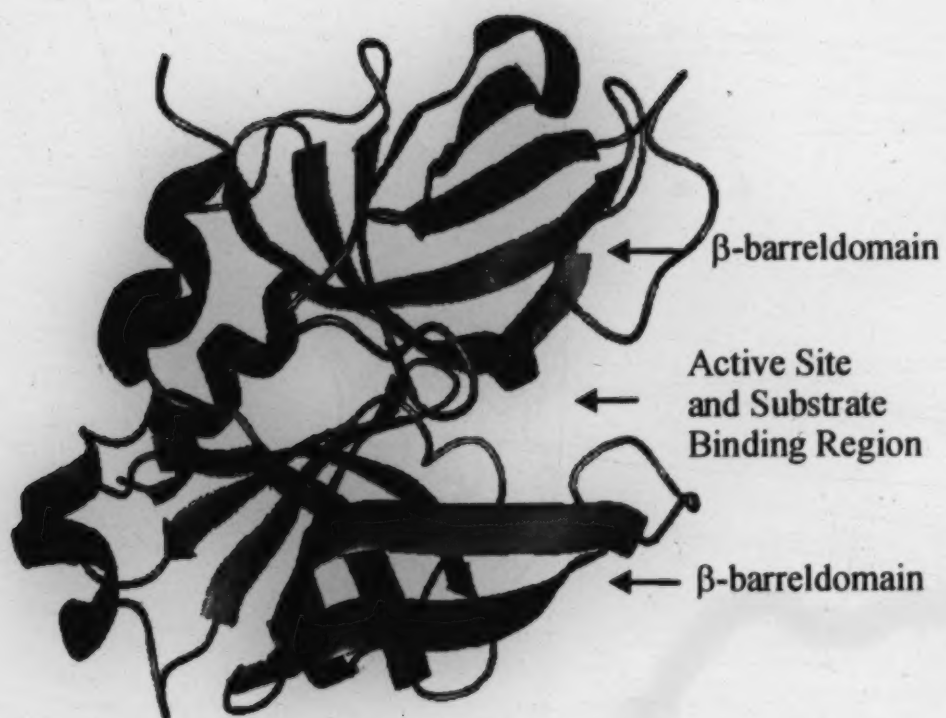
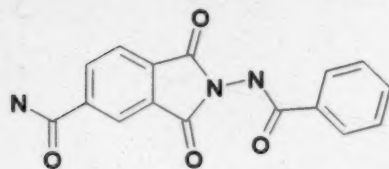


Figure 1.10. Ribbon diagram of the 3D model of HAV 3C protease as determined by X-ray crystallography (Bergmann et al., 1997).



Figure 1.11. Active site residues of HAV 3C protease.

Phthalamide inhibitors



Fluoromethylketone inhibitors

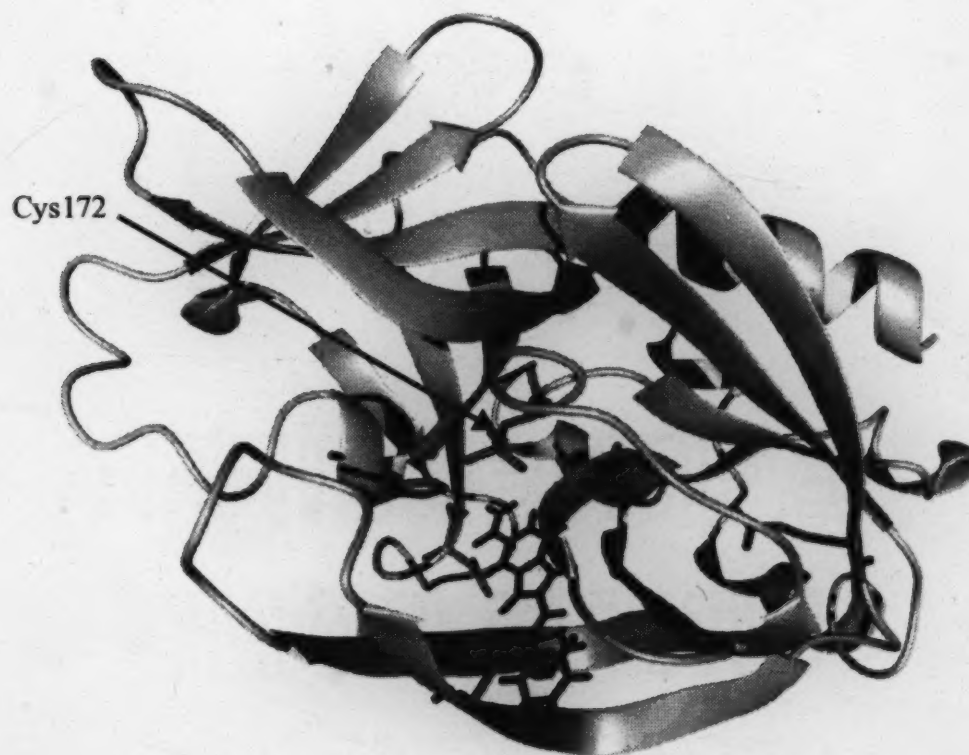
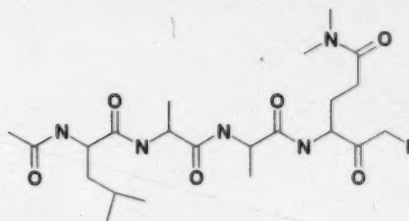


Figure 1.12. Examples of inhibitors to HAV 3C protease. The "phthalamide inhibitor" is positioned in the active site such that its aldehyde reacts with the nucleophilic thiol of Cys172.

Chapter 2

Determining solution conditions that minimize aggregation of HAV 3C protease

2.1 Introduction

A key prerequisite for determining the sequential assignments of a protein by NMR is that the protein should be soluble and non-aggregating (mono-disperse) at concentrations approaching 1 mM or higher. The increase in molecular weight associated with protein aggregation (particularly for larger proteins) will invariably lead to increases in both the rotational correlation time (τ_c) and half-height linewidths ($\Delta\nu_{1/2}$) of observed resonances. In addition, the rapid transverse relaxation (T_2) associated with such large complexes makes magnetization transfer, even via large heteronuclear couplings, a very inefficient process. Ultimately, for highly aggregated proteins, the sensitivity of NMR experiments drops dramatically and in many cases it is impossible to discern or assign resonances.

Protein aggregation may result from inter-protein electrostatic interactions, the presence of surface hydrophobic patches or the presence of mobile tails (Wagner, 1993). The extent of aggregation may depend on: pH, temperature, amount and type of salts added, addition of reducing agents such as DTT (dithiothreitol), addition of detergents such as CHAPS (3-[(3-cholamidopropyl)-dimethyl-ammonio]-1-propanesulfonate), and the addition of ligands, substrates or cofactors (Schein, 1990). Different combinations of these solution characteristics can reduce protein aggregation. However, there are no hard and fast rules to follow because each protein reacts differently to various conditions. Therefore, as a general rule, the first step in every protein NMR study should involve the optimization of the solution conditions, via thorough screening of many buffer combinations. This screening process is often termed "protein conditioning" (Wagner et al., 1993). Protein conditioning is usually geared towards finding

sample conditions for the protein that produce the narrowest linewidths possible without adversely affecting the structure of the protein.

Indications that the HAV 3C protease may be aggregating at NMR sample concentrations (~ 1 mM) were obtained after a series of NMR experiments (^1H - ^{15}N TOCSY-HSQC, the HNCACB, and the CBCACONNH) yielded spectra that were devoid of all but a handful of the expected resonances. It was postulated that efficient relaxation rates reflecting the molecular weight of an aggregated complex would account for the disappearance of the ^1H and ^{13}C signals. In further support of this premise, X-ray crystallographic experiments had earlier shown that there are two HAV 3C protease molecules present in the crystal unit cell (Bergmann et al., 1997) (Figure 2.1, p.39). Finally, previous NMR studies with the HAV 3C protease indicated protein "self-association" was a definite concern especially after the sample was left in an NMR tube for an extended period of time (Dr. Bruce Malcolm, personal communication).

To facilitate further multidimensional NMR studies with the HAV 3C protease, we decided to employ NMR spectroscopy and analytical ultracentrifugation to: (1) identify optimal sample conditions that produce the narrowest half-height linewidths, (2) determine if the protease is monomeric under these optimal sample conditions, (3) determine the maximal concentration where HAV 3C protease remains monomeric, and finally (4) determine the length of time that the protease would remain monomeric after being prepared for NMR.

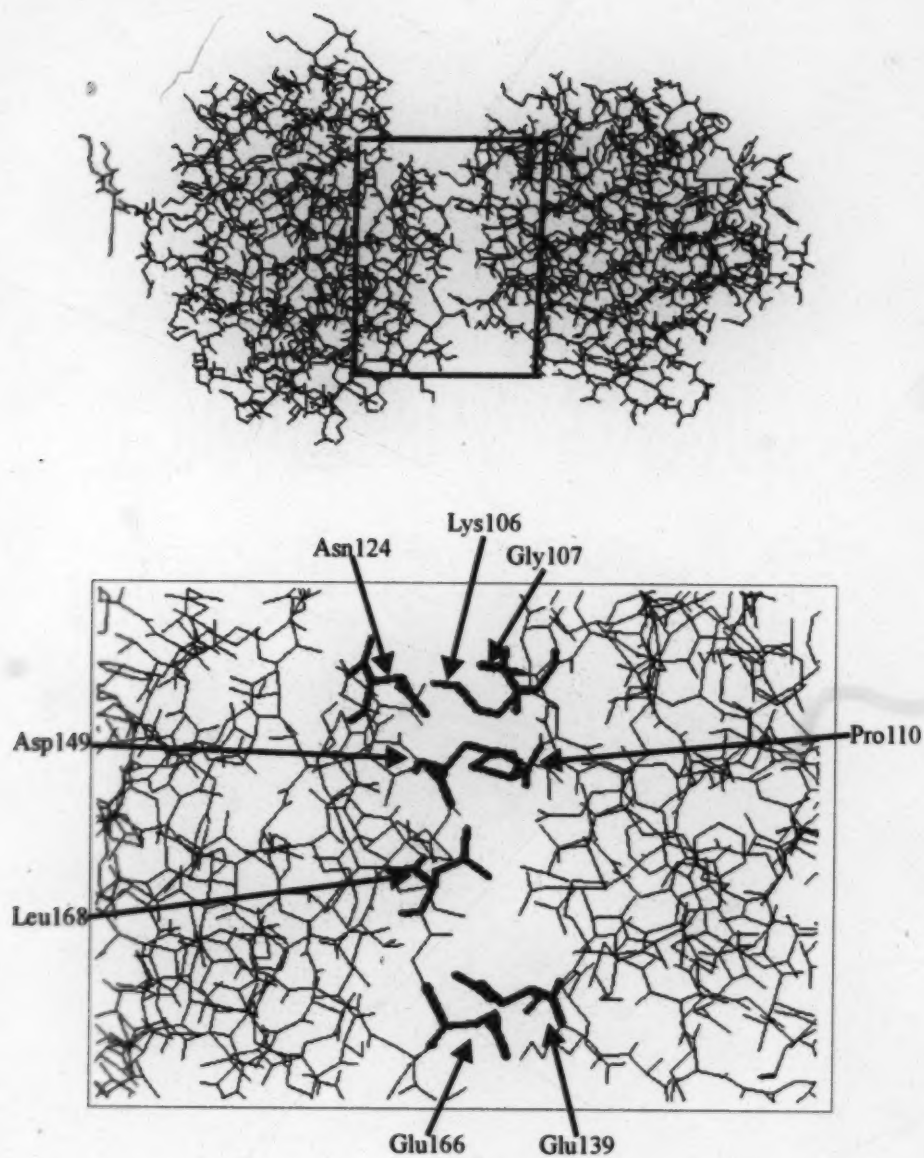


Figure 2.1. Residues involved at the dimer interface of two HAV 3C protease molecules in the unit cell as determined by X-ray crystallography (Bergmann et al., 1997).

2.2 Materials and Methods

HAV 3C Protease Expression System. The HAV 3C expression system was a generous gift from Dr. Bruce Malcolm. The expression plasmid was prepared by cloning the HAV 3C protease gene into a ~~pBR~~-based plasmid (pHAV-3CEX) containing a tac promoter to enhance expression (Malcolm et al., 1992). The HAV 3C gene was modified at position 24 (C24S) to reduce problems associated with aggregation and disulfide bridge formation noted with the wild type. The pHAV-3CEX vector was transformed into a prototrophic *E. coli* strain (MM294) that expressed high levels of HAV 3C protease upon induction with isopropyl β -D thiogalactopyranoside (IPTG).

HAV 3C Protease Expression and purification. The following protocol has been optimized from that originally presented by Malcolm et al. (1992). In general, this protocol entailed initially growing the bacterial expression system in a small (25-30 mL) "overnight" starter culture. The "overnight" culture was grown on either rich (Terrific Broth) or minimal media (M9). If sufficient bacterial growth had occurred in this overnight culture, 10 mL of this culture was used to inoculate a larger (1 L) batch of minimal media ("expression media") which was used for HAV 3C protease expression. A detailed description of this protocol follows.

Step 1. MM294 Growth and HAV 3C Protease Expression. MM294 cells, containing the HAV 3C protease gene, from -80°C freezer stock were streaked onto a LB plate containing ampicillin (150 $\mu\text{g}/\text{ml}$). The plate was incubated for ~24 hours at 37°C . A single colony was selected using a fired platinum wire and aseptically transferred into a 25-30 mL volume of media. The media was grown overnight (12-14 hours) in an incubator-shaker at 37°C and 325 rpm. 10 mL of overnight culture was used to inoculate 1 L of minimal media. Bacterial growth was monitored using optical absorbance measurements at 600 nm. When the

OD₆₀₀ reached 0.6 (~2-3 hours), IPTG was added to a final concentration of 2 mM to induce the over-expression of HAV 3C protease. Immediately after induction, the temperature of the incubator-shaker was reduced to 30°C and growth was allowed to continue for an 8 hour induction period. Bacteria were harvested by centrifuging for 10 minutes at 5,000 rpm. The supernatant was discarded and the bacterial pellet was stored at -20°C overnight.

Step 2. MM294 Lysis and HAV 3C Protease Collection. The cell pellet was resuspended in 100 mL of cell resuspension buffer. Lysozyme was added to a concentration of ~80 µg/mL lysate and the mixture was incubated on ice for approximately 30 minutes. Three continuous freeze-thaw cycles (freeze in a dry ice/ethanol mixture and thaw in 30°C water bath) were used to lyse the cells. After the freeze-thaw steps, the lysate was centrifuged for 1 hour at 18,000 rpm and the supernatant was collected. About six drops of 10% polyethylenimine was added to each 25 mL of lysate to precipitate the nucleic acids and the lysate was then centrifuged again for 30 minutes at 12,000 rpm. The supernatant was pooled and kept on ice prior to final purification.

Step 3. Purification of HAV 3C Protease. An ion-exchange chromatography column was prepared by packing a 60 mL syringe with ~ 30 mL CM Sepharose resin (Pharmacia). The column was pre-equilibrated by running ~120 mL phosphate buffer (20 mM KH₂PO₄, 0.5 mM EDTA, 1.0 mM DTT at pH 5.4) through the column using a peristaltic pump. The protein supernatant was loaded onto the column at 1mL/min. Proteins passing through the column were monitored by UV₂₈₀ absorbance using a single path monitor. After the column was loaded, the column was extensively washed with the same phosphate buffer at 1 mL/min until no further protein could be detected passing through the column. The remaining protein was then eluted with a linear gradient of NaCl (0 – 0.5 M) in the same phosphate buffer. The HAV 3C protease normally eluted at ~ 0.35 M NaCl.

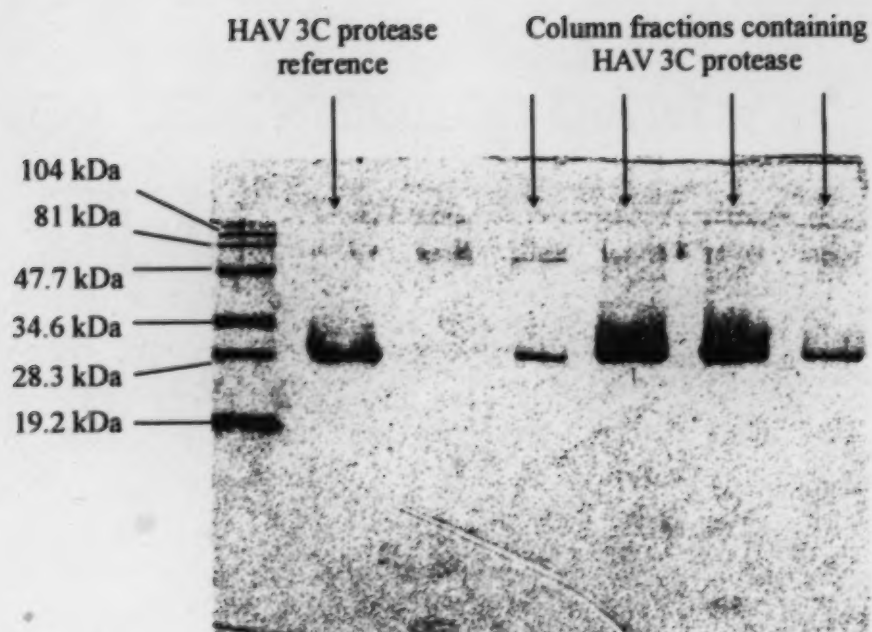


Figure 2.2. SDS-PAGE showing column fractions containing HAV 3C protease.

Step 4. HAV 3C Protease Quantification and Storage. The OD₂₈₀ was recorded for all column fractions by a UV/Visible spectrophotometer using the first fraction as a reference. All column fractions were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) according to the method of Laemmli (1970). Those fractions containing 3C protease, as judged by the appearance of a protein band similar to the HAV 3C protease reference (Figure 2.2, p.42), were pooled into a 50 mL centrifuge tube. The quantity of HAV 3C protease (mg/mL) in each fraction was calculated by dividing the fraction 'OD₂₈₀ value' by 1.2. 25 μ L of 3% w/v sodium azide was added to inhibit bacterial growth and nitrogen gas was instilled into the tube. The sealed tube was stored at 4°C. HAV 3C protease was generally found to be stable for one month under these storage conditions.

1D ¹H NMR studies. The concentration of the HAV 3C protease stock solution was determined by recording the OD₂₈₀ for the solution using a phosphate buffer (20 mM KH₂PO₄, 0.5 mM EDTA, 1.0 mM DTT at pH 5.4) as a reference. It is important to note that if pH changes were required they had to be done at this (pre-concentrated) stage as it was found changing the pH of highly concentrated protease solutions led to precipitation. Changes in pH changes were accomplished by adding drops (in the required amount) of 0.1 M KOH or 0.1 M HCl. An appropriate volume of stock solution (containing 6 mg of HAV 3C protease) was used in the preparation of each NMR sample. When 6 mg of HAV 3C protease was concentrated to 500 μ L (NMR sample volume) it would provide ~ 0.5 mM NMR samples. This was the approximate concentration of HAV 3C protease that was tested under most buffer conditions. The protease solution was concentrated to 400 – 500 μ L using a 10K centrifugal filter device (Millipore) having a MW cutoff of 10 kDa. In the series of experiments looking at salt concentration effects, the stock solution containing 350 mM NaCl was concentrated to 1.0 mL then reconstituted with 2.5 mL phosphate buffer (20 mM KH₂PO₄, 0.5 mM EDTA, 1.0 mM DTT, and pH 5.4) and then concentrated again. This yielded samples with 100 mM NaCl. To prepare samples with higher

concentrations of NaCl, the appropriate mass of NaCl was added to the already concentrated samples. Appropriate quantities of TFE, CHAPS, CaCl₂, and glycine were all added to the previously concentrated HAV 3C protease solutions in order to prepare the samples listed in Table 2.1 (p.43). Each NMR sample was prepared to a volume of 500 μ L by combining 10 μ L DSS (5 mM), 50 μ L D₂O (99%) and 440 μ L of concentrated protease in each condition set. The final step was to transfer this volume to a 5 mm NMR tube.

A 1D ¹H NMR spectrum was collected for each sample (if possible) on a Varian VXR 500 MHz NMR spectrometer equipped with a 5 mm triple resonance probe. A sweepwidth of 8000 Hz was used with 2048 transients being collected for each sample. A relaxation delay of 1.0 s was used between transients. Spectra were processed and analyzed using VNMR software version 5.1 with a line-broadening constant (lb) of 0.5.

The Val25 γ CH₃ resonance at -0.7 ppm was examined and the half-height linewidth of this resonance (Figure 2.3, p.47) was recorded for each sample. Half-height linewidths were determined using the Varian "dres" macro after suitable phasing, drift and baseline correction. The linewidths values are listed in Table 2.1.

2D ¹H-¹⁵N HMQC-J and ¹⁵N T₁/T₂-Envelope Experiments. NMR samples of the HAV 3C protease were prepared as discussed above. Four samples (A,B,C, and D) were prepared containing 1.5 mM, 1.0 mM, 0.75 mM and 0.46 mM of ¹⁵N labelled HAV 3C protease. Sample C (0.75 mM HAV 3C protease) was prepared using enzymatic assay conditions of 350 mM NaCl, 20 mM KH₂PO₄, 0.5 mM EDTA at pH 7.0. Samples A, B, and D were prepared in the ideal conditions identified by the 1D ¹H NMR linewidth studies (100 mM NaCl, 20 mM KH₂PO₄, 0.5 mM EDTA, 1.0 mM DTT at pH 5.4). Samples B and C were prepared from frozen stock (-20°C) while the other two were prepared from fresh stock stored at 4°C.

Table 2.1 Solution conditions tested in 1D ^1H NMR studies with HAV 3C protease.

Sample #	Solution Conditions	$\Delta\nu_{1/2}$ (Hz)
1	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 350 mM NaCl 0.5 mM EDTA / pH 7.0 / 25.0 °C	18
2	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 350 mM NaCl 0.5 mM EDTA / pH 7.0 / 25.0 °C / Y143G	--
3	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 350 mM NaCl 0.5 mM EDTA / pH 7.0 / 25.0 °C / Y143G with Inhibitor	--
4	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 350 mM NaCl 0.5 mM EDTA / pH 7.0 / 25.0 °C / 9.1 % (v/v) TFE	--
5	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 350 mM NaCl 0.5 mM EDTA / pH 7.0 / 25.0 °C / 4.5 % (v/v) TFE	--
6	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 350 mM NaCl 0.5 mM EDTA / pH 7.0 / 25.0 °C / 25.0 mM CHAPS	--
7	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 350 mM NaCl 0.5 mM EDTA / pH 7.0 / 25.0 °C / 7.0 mM CHAPS	19
8	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 350 mM NaCl 0.5 mM EDTA / pH 5.4 / 25.0 °C	18
9	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 100 mM NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 25.0 °C	16
10	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 350 mM NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 25.0 °C	17
11	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 450 mM NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 25.0 °C	18
12	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 750 mM NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 25.0 °C	20
13	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 1.0 M NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 25.0 °C	20
14	0.75 mM HAV 3C Protease / 20 mM KH_2PO_4 / 100 mM NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 25.0 °C	17
15	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 1.0 M NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 25.0 °C	20
16	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 1.0 M NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 30.0 °C	19
17	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 1.0 M NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 35.0 °C	19
18	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 1.0 M NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 40.0 °C	19
19	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 1.0 M NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 45.0 °C	19
20	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 100 mM NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 25.0 °C	16
21	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 100 mM NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 30.0 °C	16
22	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 100 mM NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 35.0 °C	16
23	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 100 mM NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 25.0 °C / with Inhibitor	17
24	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 100 mM NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 25.0 °C / 20 mM CaCl_2	18
25	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 100 mM NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 25.0 °C / 100 mM CaCl_2	18
26	0.5 mM HAV 3C Protease / 20 mM KH_2PO_4 / 100 mM NaCl 0.5 mM EDTA / 1.0 mM DTT / pH 5.4 / 25.0 °C / 9.1 % (v/v) Saturated Glycine Solution	--

2D ^{15}N - ^1H HMQC-J experiment. The 2D ^{15}N - ^1H HMQC-J experiment (Kay et al., 1989; Kay and Bax, 1990) was used to determine $^3J_{\text{HNH}\alpha}$ coupling constants. This experiment differs from the 2D ^{15}N - ^1H HMQC in that it utilizes a $90^\circ\gamma$ purge pulse (Frey et al., 1985) to eliminate the dispersive contributions to lineshape that are associated with J modulation present during the acquisition time, t_1 . A key to this experiment was collecting a high number of transients to allow for good spectral resolution. In the ^{15}N - ^1H HMQC-J experiment (Figure 2.4, p.48), magnetization is generated on $^1\text{H}_\text{N}$ [Step 1]. An HMQC pulse sequence transfers magnetization to ^{15}N via the $^1J_{\text{NH}}$ (~ 90 Hz) coupling [Step 2]. ^{15}N chemical shifts are recorded over the acquisition time t_1 [Step 3]. A second HMQC pulse sequence transfers magnetization back to $^1\text{H}_\text{N}$ [Step 4] and the proton chemical shift is collected over the second acquisition time t_2 [Step 5]. Fourier transformation of the resultant FID yields a frequency domain map showing all $^1\text{H}_\text{N}$ and ^{15}N correlations.

^{15}N T_1/T_2 -Envelope Experiments. The ^{15}N T_1 and T_2 relaxation data for the amide envelope were obtained by recording a series of 1D ^{15}N inverse detected NMR spectra for a series of T_1 and T_2 delays (Farrow et al., 1994). In these experiments, an INEPT sequence was used to transfer the magnetization from $^1\text{H}_\text{N}$ to the directly attached ^{15}N . The ^{15}N magnetization was refocused in either the z plane (for T_1 -envelope) or the xy plane (for T_2 -envelope). Then the ^{15}N magnetization was allowed to evolve for a period of time (T_{xd}) specified by

$$T_{1d} = \text{ncyc} \{ \text{PW\_SHPSS} + (2 \times 2.5 \times 10^{-3}) \} \quad [2.1]$$

$$T_{2d} = \text{ncyc} \{ (32 \times \text{PWN}) + (32 \times 450 \times 10^{-6}) \} \quad [2.2]$$

where ncyc was the number of cycles used (1 to 10), PW_SHPSS (cosine-modulated 180° ^1H pulse) = 550×10^{-6} s, and PWN (90° ^{15}N pulse) = 59×10^{-6} s. This corresponded to T_{1d} values of 0.0056 s, 0.0112 s, 0.0168 s, 0.0224 s, 0.0280 s, 0.0336 s, 0.0392 s, 0.0448 s, 0.0504 s and 0.0560 s. T_{2d} values were 0.0163 s, 0.0326 s, 0.0489 s, 0.0652 s, 0.0815 s, 0.0978 s, 0.1141 s, 0.1304 s, 0.1467 s and 0.1630 s. After incrementally increasing the amount of time, a

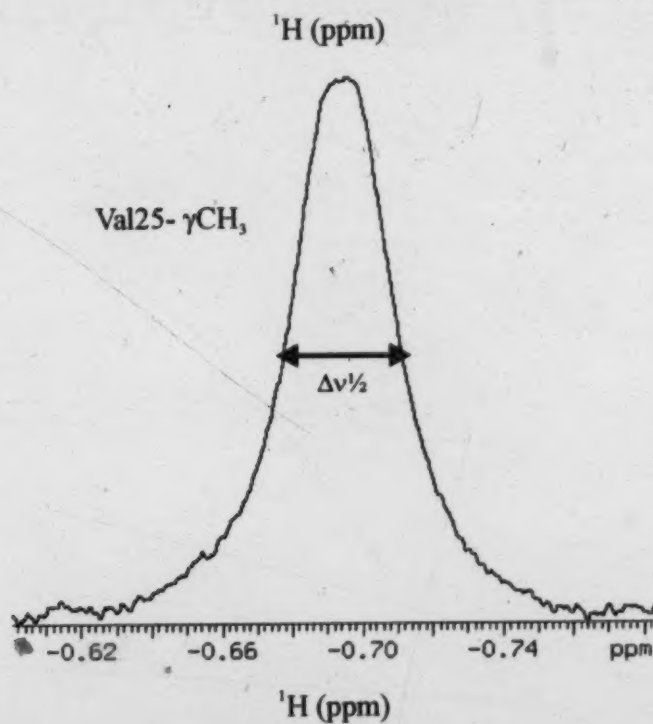
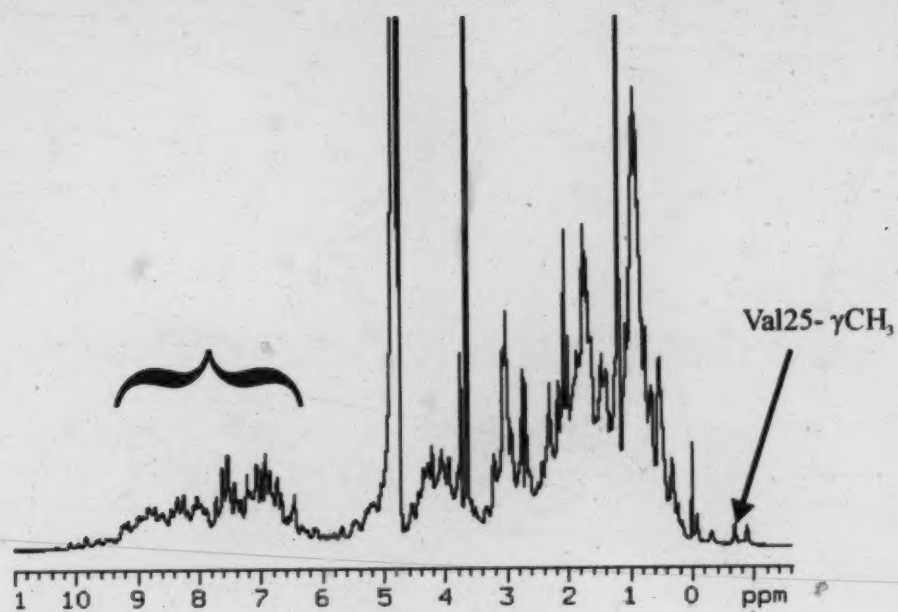


Figure 2.3. Comparing the effect of NMR buffer conditions on the HAV 3C protease by using 1D ^1H NMR studies.

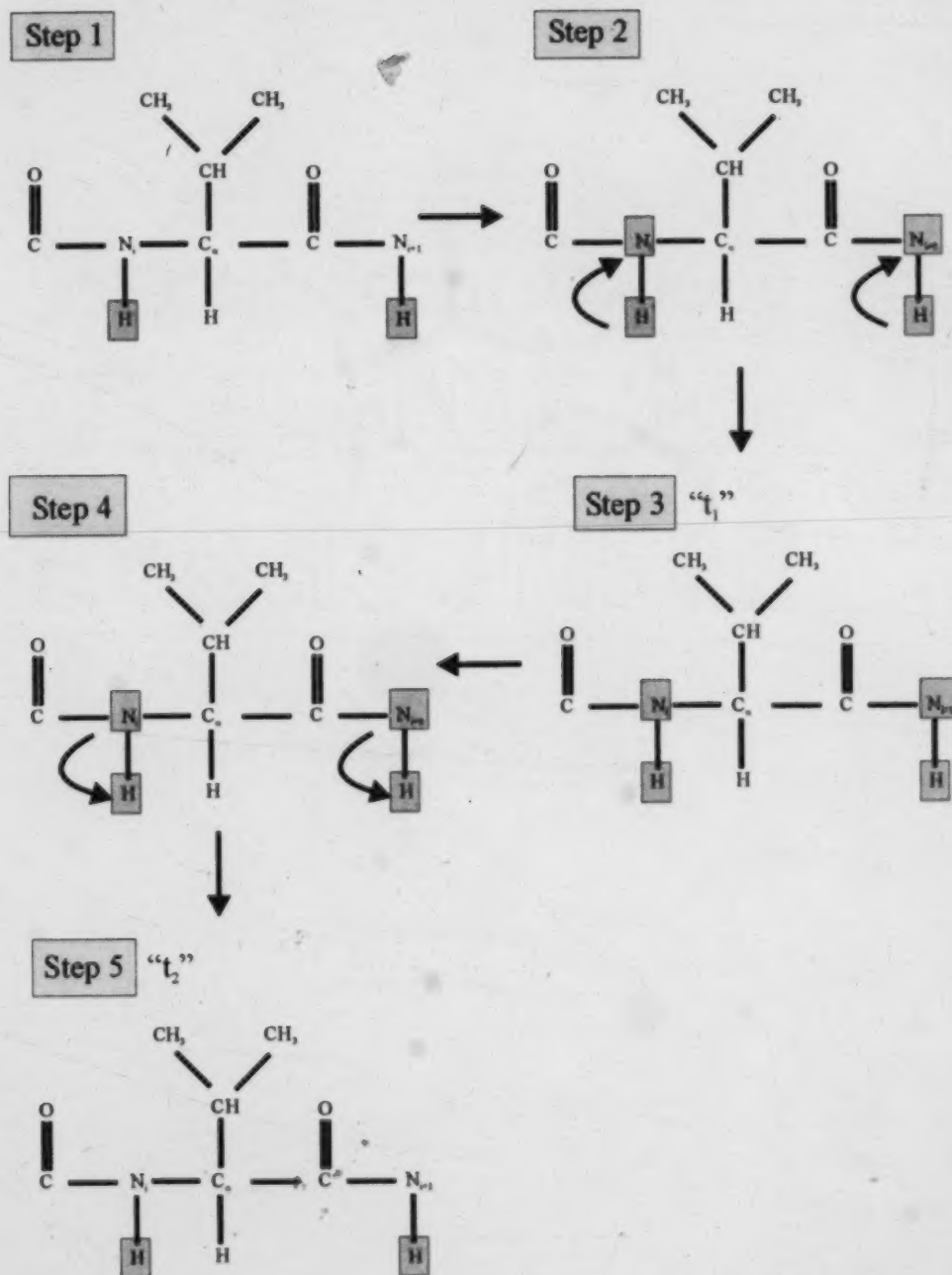


Figure 2.4. Outline of magnetization transfer in the HMQC-J experiment.

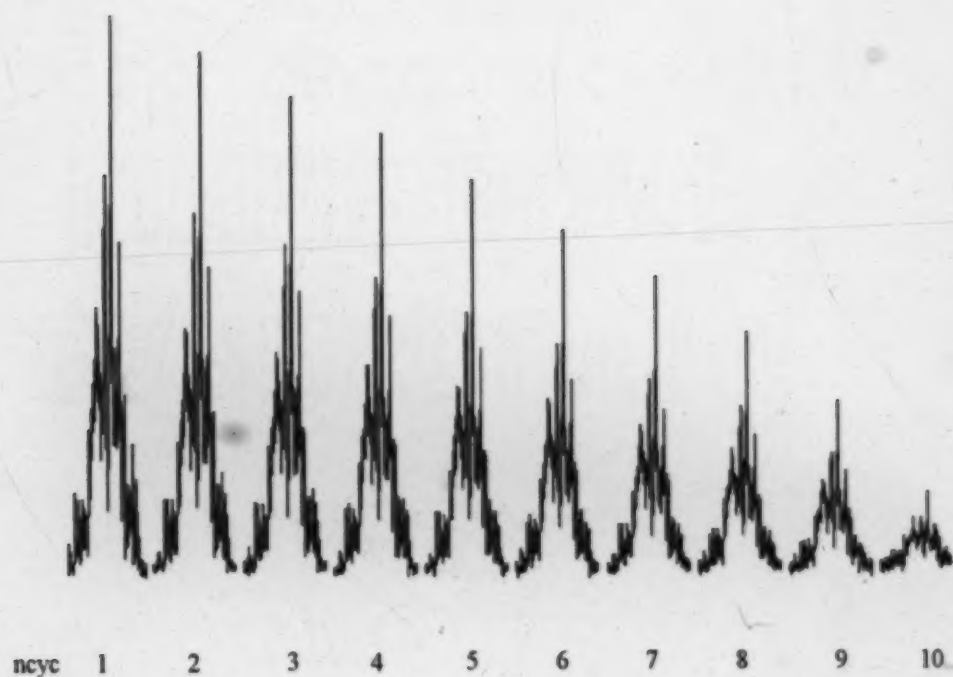


Figure 2.5. A typical result for ^{15}N T_1 - and T_2 -envelope experiments. The amide envelope is shown decreasing logarithmically as ncyc increases from 1 to 10 (successive increases in T_{1d} or T_{2d}).

reverse INEPT pulse sequence was used to transfer magnetization back to the proton for signal acquisition.

$^{15}\text{N-T}_1$ and $^{15}\text{N-T}_2$ values were obtained by integrating the area under the entire $^1\text{H}_\text{N}$ envelope for each cycle in the experiment (Figure 2.5, p.49). As magnetization evolved on the ^{15}N during the delay, the magnetization would also undergo relaxation. A longer delay corresponded to a greater period of relaxation and a further decay of magnetization. Therefore, each cycle would transfer less and less magnetization back to the proton for detection. Consequently, the integral of the amide envelope would decrease in an exponential fashion throughout the experiment. The ' T_{1d} or T_{2d} - normalized integral value' values (Table 2.2) were entered into the program "xcrvfit" (performed at PENCE, University of Alberta) which calculated values for $^{15}\text{N-T}_1$ and $^{15}\text{N-T}_2$. Xcrvfit (Robert Boyko, unpublished) calculates these values using iterative numerical methods to determine the best-fit curve to the predicted exponential.

Table 2.2. Example of normalized "amide-envelope" integral data used to calculate T_1 and T_2 values using 'xcrvfit'. This data set was collected from a 0.75 mM HAV 3C protease sample in 350 mM NaCl, 20 mM KH_2PO_4 , 0.5 mM EDTA and pH 7.0.

ncyc	$^{15}\text{N-T}_1$ Envelope		$^{15}\text{N-T}_2$ Envelope	
	T_{1d} (s)	^1H N Integral	T_{2d} (s)	^1H N Integral
1	0.0056	1.00	0.0163	1.00
2	0.0112	0.93	0.0326	0.64
3	0.0168	0.86	0.0489	0.51
4	0.0224	0.78	0.0652	0.36
5	0.0280	0.71	0.0815	0.32
6	0.0336	0.63	0.0978	0.23
7	0.0392	0.54	0.1141	0.20
8	0.0448	0.45	0.1304	0.14
9	0.0504	0.33	0.1467	0.12
10	0.0560	0.18	0.1630	0.08

Experimental Set Up. Dr. Carolyn Slupsky and Dr. Stephane Gange performed the NMR experiments. ^1H - ^{15}N HMQC-J and ^{15}N T_1 - and T_2 -envelope spectra were recorded at 25°C on a Varian Unity 600 spectrometer equipped with a triple resonance probe. ^1H - ^{15}N HMQC-J experiments were performed with a ^{15}N sweepwidth of 2000 Hz and a ^1H sweepwidth of 6000 Hz. A total of 2048 complex points were collected along the t_2 domain (^1H) and 640 increments along the t_1 domain (^{15}N). The relaxation delays for these experiments were 1.1 s and the refocusing delay was set at 0.0024 s. ^{15}N T_1 - and T_2 -envelope experiments were performed with a ^{15}N sweepwidth of 1500 Hz and a ^1H sweepwidth of 8000 Hz. A total of 1216 complex points were collected in the t_2 domain (^1H) and 2 increments in the t_1 domain (^{15}N). Relaxation delays of 2.0 s and 2.5 s were employed for ^{15}N T_1 - and T_2 -envelope experiments respectively.

Analytical Ultracentrifugation studies. Two solutions of HAV 3C protease (5 mg/mL) were prepared by concentrating HAV 3C protease stock solution (1-3 mg/mL) using a 10K centrifugal filter device (Millipore) with a MW cutoff of 10 kDa. The first buffer (the original storage conditions or enzymatic assay conditions) contained 350 mM NaCl, 20 mM KH_2PO_4 , 0.5 mM EDTA at pH 7.0. The second buffer (optimized storage conditions minus 250 mM NaCl) contained 100 mM NaCl, 20 mM KH_2PO_4 , 0.5 mM EDTA, 1.0 mM DTT, at pH 5.4. In the second buffer, salt was removed by concentrating the stock solution, reconstituting with an appropriate volume of phosphate buffer (20 mM KH_2PO_4 , 0.5 mM EDTA, 1.0 mM DTT, and pH 5.4), and then concentrating again to 5 mg/mL HAV 3C protease.

Les Hicks, of Dr. Cyril Kay's laboratory, performed the sedimentation equilibrium experiments. They were carried out at 20°C using a Beckman XLI Analytical Ultracentrifuge and interference optics following procedures outlined in the XLI Instruction Manual. 110 μL aliquots of HAV 3C protease solution were loaded into 6-sector CFE (charcoal filled epon) sample cells, allowing three

concentrations (5.48 mg/mL, 3.28 mg/mL, and 1.1 mg/mL) of sample to be run simultaneously. Runs were performed at a 16,000 and 20,000 rpm and were continued until there was no significant difference in scans taken approximately three hours apart. The sedimentation equilibrium data was evaluated using a nonlinear least-squares curve-fitting algorithm (Johnson et al., 1981) contained in the NonLin analysis package.

HAV 3C Protease NMR Sample Stability Study. HAV 3C protease stock (2 mg/mL) was stored at 4°C in the optimized post-column buffer (350 mM NaCl, 20 mM KH₂PO₄, 0.5 mM EDTA and 1 mM DTT at pH 5.4). After a month of storage, this stock was used to prepare two 0.63 mM HAV 3C protease NMR samples in a buffer containing 100 mM NaCl, 20 mM KH₂PO₄, 0.5 mM EDTA, 1.0 mM DTT at pH 5.4. A 1D ¹H NMR spectrum was collected for each of these samples on days 1, 4, 7, 9, and 27. In between collection days, one sample was stored at 4°C and the other at room temperature (25°C). All 1D ¹H NMR spectra were recorded and analyzed as described previously.

2.3 Results and Discussion

Self-association measurements of HAV 3C protease. While ¹H-¹⁵N HSQC and ¹H-¹⁵N TOCSY-HSQC spectra of the ¹⁵N labelled HAV 3C protease generally yielded good spectra with well-resolved resonances, later experiments conducted on doubly labelled (¹³C/¹⁵N) protease generally gave disappointing results. In particular, the HNCACB and CBCACONNH triple resonance experiments displayed only a few intense peaks that were presumably from the C-terminal residues.

Initially, a low level of isotopic substitution was thought to be the cause of these poor results. However, subsequent electrospray mass spectrometry analysis revealed that isotopic incorporation was >95% for both ¹³C and ¹⁵N. This result is

presented in Chapter 3 that discusses isotopic incorporation into HAV 3C protease. As low labelling efficiency could be ruled out, an alternative explanation was that the HAV 3C protease was aggregating into a polymeric state. Aggregation or increased apparent MW could account for the poor quality of those spectra that depended on multiple polarization transfers (HNCACB etc.) while at the same time it could explain the relatively good quality HSQC spectra (where only two polarization transfers are used) observed for the ^{15}N labelled material.

Another source of information that indicated that the HAV 3C protease may have a tendency to associate came from the analysis of the ^1H - ^{15}N HMQC-J experiments. ^1H linewidths in the range of 18 to 30 Hz, were measured from a ^1H - ^{15}N HMQC-J spectrum collected from a 0.75 mM HAV 3C protease sample in 350 mM NaCl, 20 mM KH_2PO_4 , 0.5mM EDTA at pH 7.0 (high salt/high pH). Three-bond $\text{H}_\text{N}\text{H}_\alpha$ coupling constants ($J_{\text{HNH}\alpha}$) can be calculated using the following formula:

$$J_{\text{HNH}\alpha} = 0.50(\Delta\nu_{1/2}) - \text{MW}/10\ 000 \quad [2.3]$$

Where $\Delta\nu_{1/2}$ is the ^1H half height linewidth and MW is the molecular weight (Wishart and Wang, 1998). Using this formula and the measured ^1H linewidths, $\text{H}_\text{N}\text{H}_\alpha$ coupling constants were calculated to be in the range of 6.77 – 13.10 Hz when a MW of 23.87 kDa (monomeric species) was used. If the same formula was used with a MW of 47.64 kDa (dimeric species), coupling constants in the range of 4.37 – 10.7 Hz were calculated. This latter range is more typical for proteins (between 3 Hz and 10 Hz) and therefore this result suggested dimerization was likely occurring.

To further ascertain if HAV 3C protease aggregation was the primary cause of our poor NMR results, we performed a sedimentation equilibrium study using analytical ultracentrifugation. This study clearly showed that the protease was associating in the original high salt/high pH conditions (5mg/mL HAV 3C

protease, 350 mM NaCl, 20 mM KH_2PO_4 , 0.5 mM EDTA, pH 7.0). The results of this study, displayed in Figure 2.6 (p.55), show that the solution behavior of this protease doesn't fit to a single species model. The absence of a 'shotgun appearance' for the residuals is a key indicator of a poor fit. In addition, the standard deviation is 0.087 when it should be less than 0.02 for a good fit. Many different 'species' models (monomer, monomer-dimer, monomer-trimer, monomer-dimer-trimer) were tested and it was found that the "single-species" (monomer) model yielded the best result. Using this model, the apparent molecular weight was determined to be 34,547 daltons whereas the theoretical mass of HAV 3C protease was 23,876.59 daltons based on the amino acid sequence and the natural abundance of all isotopes. This result indicated that the HAV 3C protease was "self-associating" to some degree, even at the low concentrations ($\sim 0.04 - 0.30$ mM) used for ultracentrifugation. Consequently, a large scale screening of NMR sample conditions was performed to determine optimal NMR sample conditions that might reduce the self-association tendency of HAV 3C protease.

Identification of optimal NMR sample conditions. In order to do determine optimal conditions for NMR analysis, two different HAV 3C protease mutants and nearly 30 different solution conditions were studied. These conditions included different temperatures, varying salt concentrations, varying pH, and the addition of organic solvents and detergents. The actual solution conditions for each NMR sample tested are listed in Table 2.1. Qualitative information about the presence of aggregation can be obtained by monitoring the broadening of NMR linewidths (Akiyoshi et al., 1993). Even though the analysis of the half-height linewidth ($\Delta\nu_{1/2}$) does not provide easily quantifiable information about the size of the protein aggregates, the narrowing of linewidths qualitatively corresponds to longer T_2 values and/or a decreases in molecular weight. Table 2.1 outlines these solution studies and the recorded half-height linewidth ($\Delta\nu_{1/2}$) of the upfield Val25 CH_3 resonance (at -0.7 ppm) for each sample.

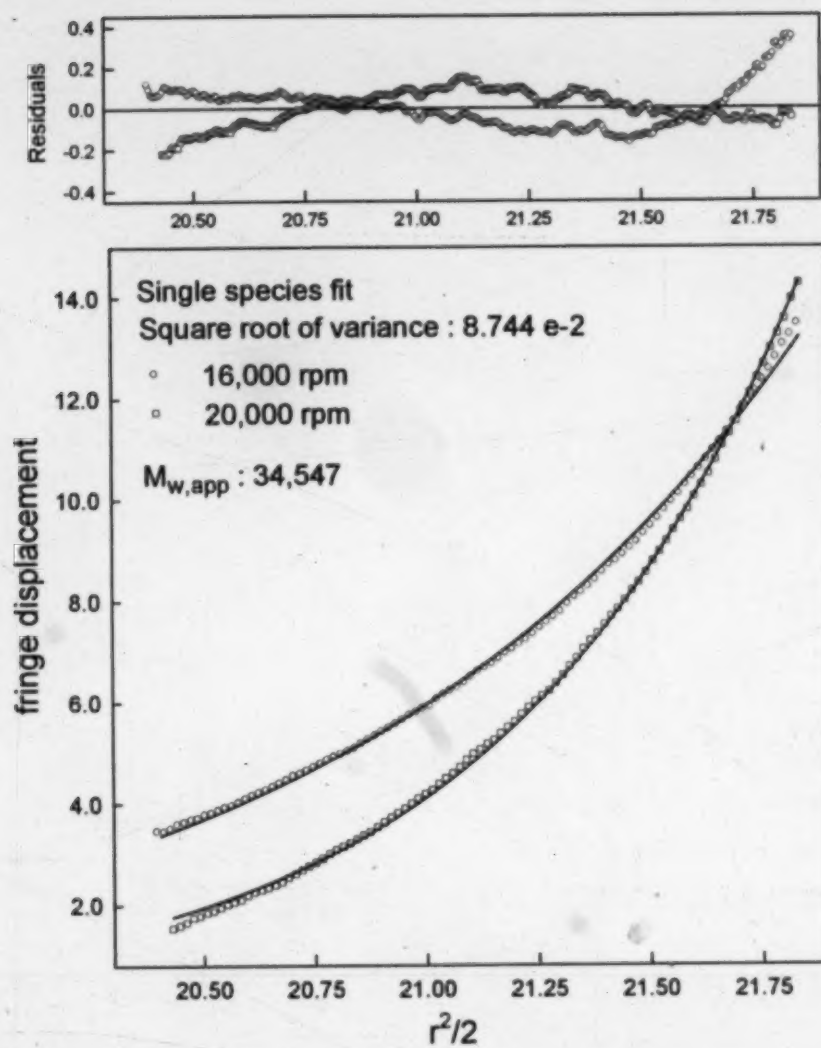


Figure 2.3: Sedimentation Analysis of HAV 3C protease in 350 mM NaCl, 20 mM KH_2PO_4 , 0.5 mM EDTA, and pH 7.0.

Frozen vs. unfrozen HAV 3C protease stock solutions. During the initial analysis of these 1D ^1H NMR spectra, it was observed that better resolved spectra were obtained when HAV 3C protease samples were prepared from pooled protease stock that had not been previously frozen. Figure 2.7 (p.57) shows how the amide region of the unfrozen spectrum displayed sharper peaks than the sample that had been previously frozen. "Freeze-thaw" induced aggregation has been previously observed with other proteins such as recombinant human factor XIII. This was attributed to interfacial adsorption and denaturation at the air-liquid and ice-liquid interfaces (Kreilgaard et al., 1998). With respect to the HAV 3C protease, it is likely that as ice crystals formed during the freezing process, protease molecules were forced too close together (Martin, 1993). Upon thawing, precipitation was not observed but it could have been that some degree of freeze-induced or interfacial denaturation had taken place. This "freeze-thaw" induced denaturation, accompanied with the proximity of other HAV 3C protease molecules in the frozen state, likely led to an increased tendency to self-associate. After this observation was made, all pooled protease stock was stored in the refrigerator at 4°C instead of freezing it at -20°C.

Preventing hydrophobic interactions between HAV 3C protease molecules. If protein aggregation arises from the interaction of hydrophobic interfaces between monomers, site-directed mutagenesis at the hydrophobic regions can be used to alleviate the aggregation problem (Schein, 1990). This method has been used successfully for the solution structure determination of the insulin monomer (Kaarsholm and Ludvigsen, 1994). To investigate whether site-directed mutants of HAV 3C protease could be prepared (or found) that would have a reduced tendency to aggregate, we analyzed the X-ray structure of HAV 3C protease (Bergmann et al., 1997). Interestingly this structure shows two molecules interacting as a dimer (Figure 2.1). To analyze the dimer interface, a program called VADAR was used (Wishart, 1994). VADAR provides information on accessible surface area and packing volumes and can be used to quantify the interactions between amino acids in domains or interfaces. Structural analysis was

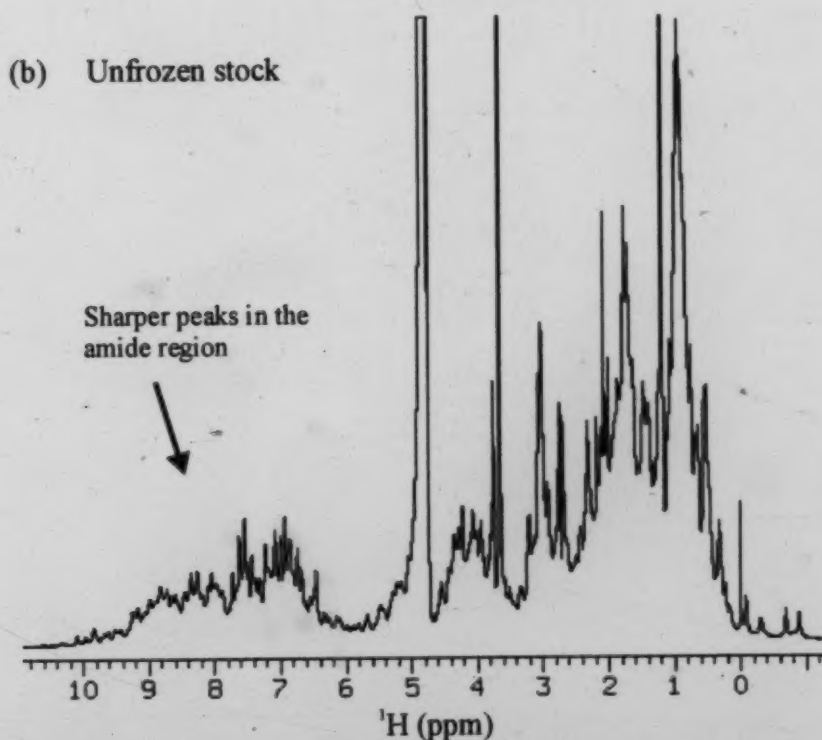
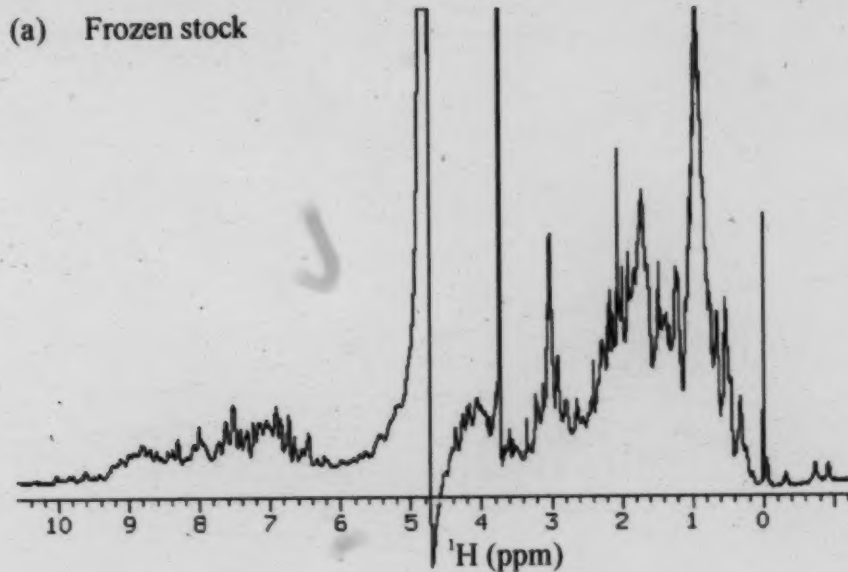


Figure 2.7. Comparison of 1D ^1H NMR spectra collected from 0.5 mM HAV 3C protease sample (20mM KH_2PO_4 , 350 mM NaCl, 0.5 mM EDTA at pH 7.0) prepared with (a) frozen and (b) un-frozen stock solutions at 25 $^\circ\text{C}$ on a Varian VXR 500 MHz NMR Spectrometer.

performed on HAV 3C protease's PDB file (1HAV) to determine which residues appear to be interacting at the interface. By running VADAR on the crystallographic dimer and then taking out the coordinates of one of the species and rerunning the VADAR surface area analysis, we were able to determine which amino acids are most likely involved in the dimeric interfaced interactions. The residues that showed the greatest change in surface area were Asn124, Glu166, Leu168, Gly107, Pro110, and Glu139. For a study involving site-directed mutagenesis, it would appear that replacing the most hydrophobic residue, in this case Leu168, with a hydrophilic residue type (perhaps lysine) might potentially reduce the tendency to dimerize. Additionally, because electrostatic interactions were suspected for the HAV 3C protease, it may have been beneficial to substitute Glu166 (or Glu139) with an alanine in order to prevent this interaction. While the Leu168Lys (L168K), Glu166Ala (E166A) or Glu139Ala (E139A) mutants were not available to investigate this hypothesis, we were able to analyze another mutant (Y143G) that had been previously found to exhibit "better than average" spectra (Dr. W. Bachovchin, personal communication).

Mutants of HAV 3C protease. Spectra from the Y143G mutant had unusual lineshapes that appeared very narrow in the upper portions of the resonance but broad in the lower portions (Figure 2.8, p.60). This baseline broadening could be due to the final structure being composed of two slightly different molecular conformations. An additional problem arose with this mutant as well. In particular, the HAV 3C Y143G mutant protease could only be expressed and purified in cells grown in rich (TB) media and not in minimal media which is required for isotopic labelling. For reasons of the baseline broadening and expression problems, we decided not to pursue further NMR studies with the Y143G mutant and, instead, focused on improving conditions for the original HAV 3C protein.

Effect of solvents or detergents. Another method of treating aggregation is to add detergents or organic solvents (Schein, 1990). Detergents are amphiphilic compounds that lower the surface tension of water and bind to hydrophobic areas

on proteins thereby preventing aggregation. CHAPS is a detergent that has been shown to be quite effective in reducing aggregation and it was successfully used in the multidimensional studies of calcineurin B (Anglister, et al., 1993). Two HAV 3C protease samples were prepared with 25.0 and 7.0 mM concentration of CHAPS. In both cases, the C24S mutant showed no narrowing of half-height linewidth. Denaturing solvents such as trifluoroethanol (TFE) have also assisted in reducing aggregation by providing a competitive interaction for the intramolecular hydrophobic interactions responsible for stable tertiary structure (Schein, 1990). TFE was successfully used for the NMR study of troponin C (Slupsky, et al., 1995). However, in our hands, any amount of TFE added to HAV 3C protease caused immediate precipitation of the protein. The fact that CHAPS did not lead to half-height linewidth narrowing suggests that electrostatic interactions, rather than hydrophobic interactions, may be the predominant factor leading to self-association of the HAV 3C protease.

► **Effect of pH.** To determine the effect of pH on HAV 3C protease spectra, we also prepared a sample at pH 5.4. While there was no appreciable narrowing of linewidth at the lower pH, the amide region of spectra appeared to have sharper peaks. In addition, a ^1H - ^{15}N HSQC spectra of a doubly labelled HAV 3C protease (^{13}C labelled Glycine and ^{15}N labelled Leucine) taken at pH 5.4 was compared to a similar spectra collected at pH 7.0 of the protease specifically labelled at leucine. The spectra collected at lower pH appeared to be better resolved than those spectra recorded at higher pH. The lower pH likely increases the amide signal due to the slower hydrogen exchange rate at acidic pHs'. Based on these results, It was decided that all further experiments would be performed at pH 5.4.

Effect of salt concentration. Salt concentrations played a critical role in defining the optimal solution conditions for HAV 3C protease. In Table 2.1 (p. 45), it is noted that decreasing the NaCl concentration from 1.0 M to 0.1 M narrowed the half-height linewidth from 20 Hz to 16 Hz. The reason salt concentration had such

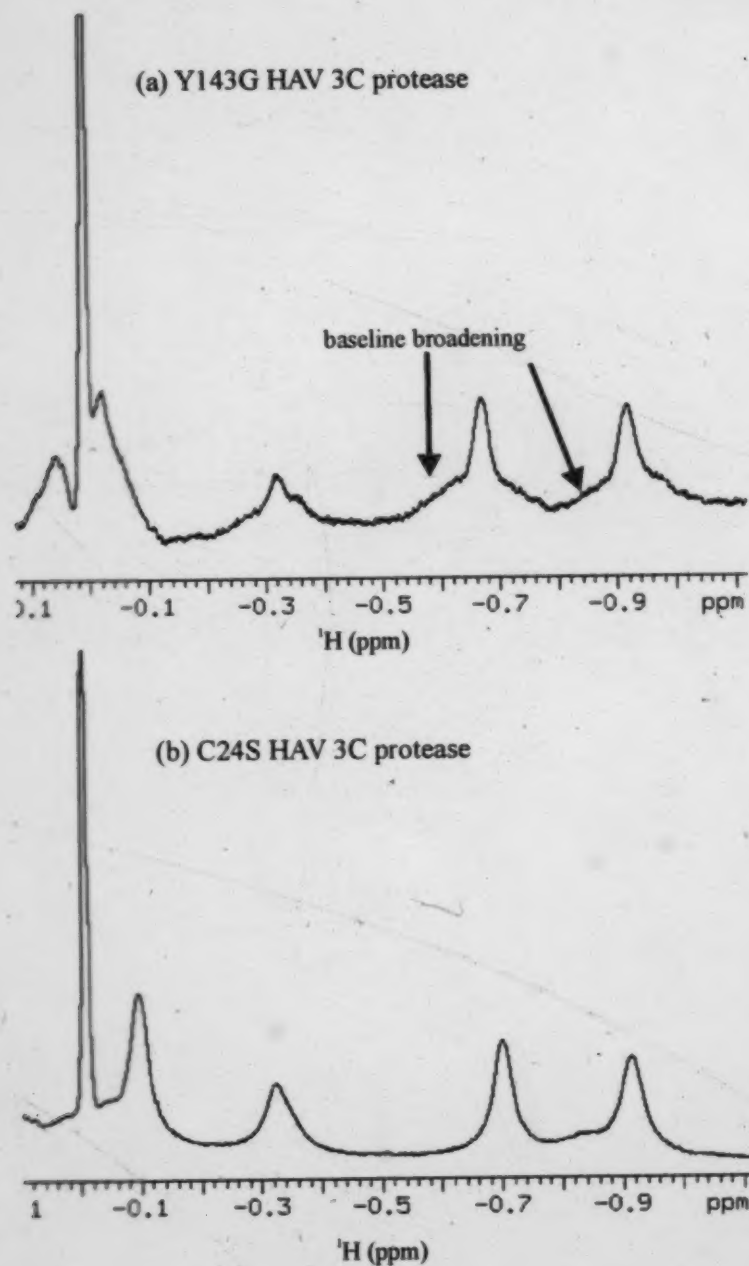


Figure 2.8. Comparison of 1D ¹H NMR spectra collected from 0.5 mM (a) Y143G and (b) C24S mutants of HAV 3C protease samples (20 mM KH₂PO₄, 350 mM NaCl; 0.5 mM EDTA at pH 7.0) at 25°C on a Varian VXR 500 Mhz NMR Spectrometer.

an impact on the half-height linewidth can potentially be explained by the DLVO (Deryagin-Landau-Verwey-Overbeek) theory of colloids (De Young et al., 1993). This theory relates the attractive van der Waals and repulsive electrostatic forces as a function of intermolecular distance (Figure 2.9, p.62). As the protease concentration increases, the distance between molecules would decrease. In order to retain electrostatic repulsive forces between molecules at high concentration (smaller distances between molecules), this theory states that a low salt concentration is required. At high salt concentrations, the electrostatic charges on the protease are shielded from each other and the van der Waals (attractive) forces dominate. This results in an attraction between molecules that ultimately leads to aggregation. Lowering the salt concentration ensures that electrostatic repulsive forces would remain between molecules thereby preventing a polymeric state from arising. Although the linewidth change from 20 Hz to 16 Hz does not by itself mean that an aggregated state has been transformed into a monomeric state, it does suggest that the protease is behaving better in NMR samples with low salt concentrations.

As can be seen in Table 2.1 (p. 45), certain solution properties didn't appear to affect the half-height linewidth. These factors were temperature and the addition of DTT and CaCl_2 to the solution. However, the temperature studies did teach us something about the stability of the protein. Increasing the temperature generally led to more rapid precipitation. Performing NMR experiments above 25°C would typically cause the protein to precipitate within hours (depending on the temperature). The highest temperature that could be tolerated for extended periods of time (10 days) was 25°C . While DTT didn't appear to have an effect on linewidth it was added to the samples to prevent disulfide bridge formation. In the C24S HAV 3C protease, there is still the active site Cys172 that could potentially lead to the formation of intermolecular disulfide linkages. Erring on the side of caution, we decided to add DTT to all samples to keep the Cys172 in the reduced form.

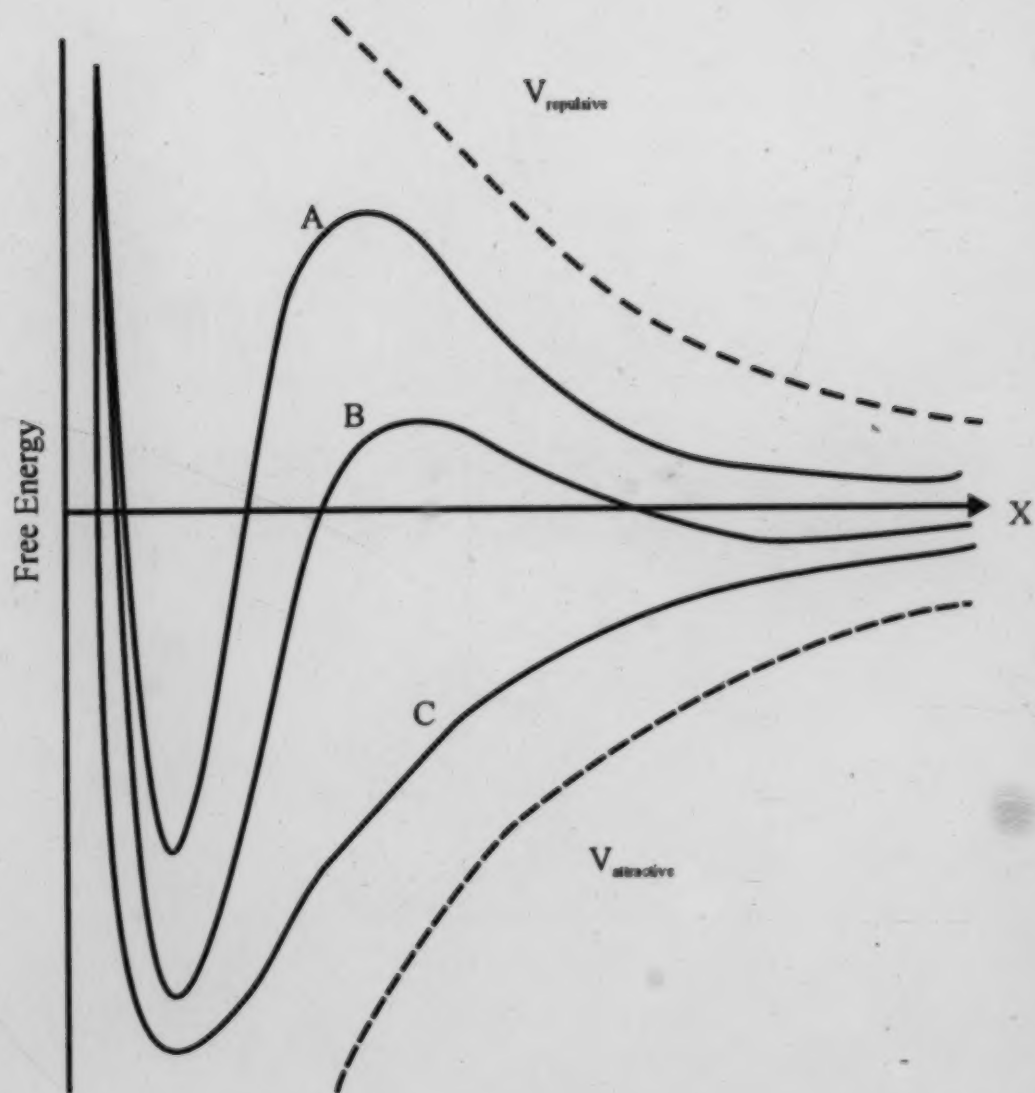


Figure 2.9. DLVO Theory. Van der Waals electrostatic forces ($V_{\text{attractive}}$ and $V_{\text{repulsive}}$) are plotted as a function of intermolecular distance (x). (a) Low salt, (b) medium salt and (c) high salt.

In conclusion, the 1D ^1H NMR half height line width studies indicate that the optimum NMR sample conditions for the HAV 3C protease are: 100 mM NaCl, 20 mM KH_2PO_4 , 0.5 mM EDTA, 1.0 mM DTT, at pH 5.4

Confirmation of a HAV 3C protease monomer. Once optimal solution conditions were determined, analytical ultracentrifugation of HAV 3C protease under the new conditions was performed to determine if the protein was monomeric. The results of the second ultracentrifugation study are presented in Figure 2.10 (p.64). Under these low salt, low pH conditions, HAV 3C protease clearly adhered to a single species fit. The fit in this study is 'good' because of the shotgun appearance of the residuals and the standard deviation is just 0.017. The results show an apparent molecular weight of 24,719 daltons which is significantly closer to the theoretical molecular weight of 23,876.59 daltons based on the amino acid sequence and the natural abundance of all isotopes. Fitting the data to a monomer-dimer species model, a dissociation constant (K_d) of 13 mM was determined. This means that at 13 mM, under these solution conditions, 50 per cent of the HAV 3C protease molecules would be dimerized.

^{15}N - T_1 and ^{15}N - T_2 envelope studies (Farrow et al., 1995) were also performed to confirm the results of analytical ultracentrifugation and to investigate the maximal allowable concentration for the HAV 3C protease. Table 2.3 shows the backbone ^{15}N - T_1/T_2 results for one sample (C) in the high salt, high pH conditions and three samples (A,B,D) collected under optimal (low salt, low pH) conditions. Interestingly, these data agree very well with those from the analytical ultracentrifugation studies. Both techniques indicate that the molecular weight has decreased from ~ 32 to ~24 kDa (approximately monomeric) when the protein was moved from high salt, high pH conditions to low salt, low pH conditions. If the protease was existing as a dimer in the original NMR sample conditions the estimated molecular weight would have been ~ 48 kDa. This implies that the protease was not fully dimerized in the high salt, high pH conditions but a certain percentage of self-association was evident. This percentage was ~ 45% and

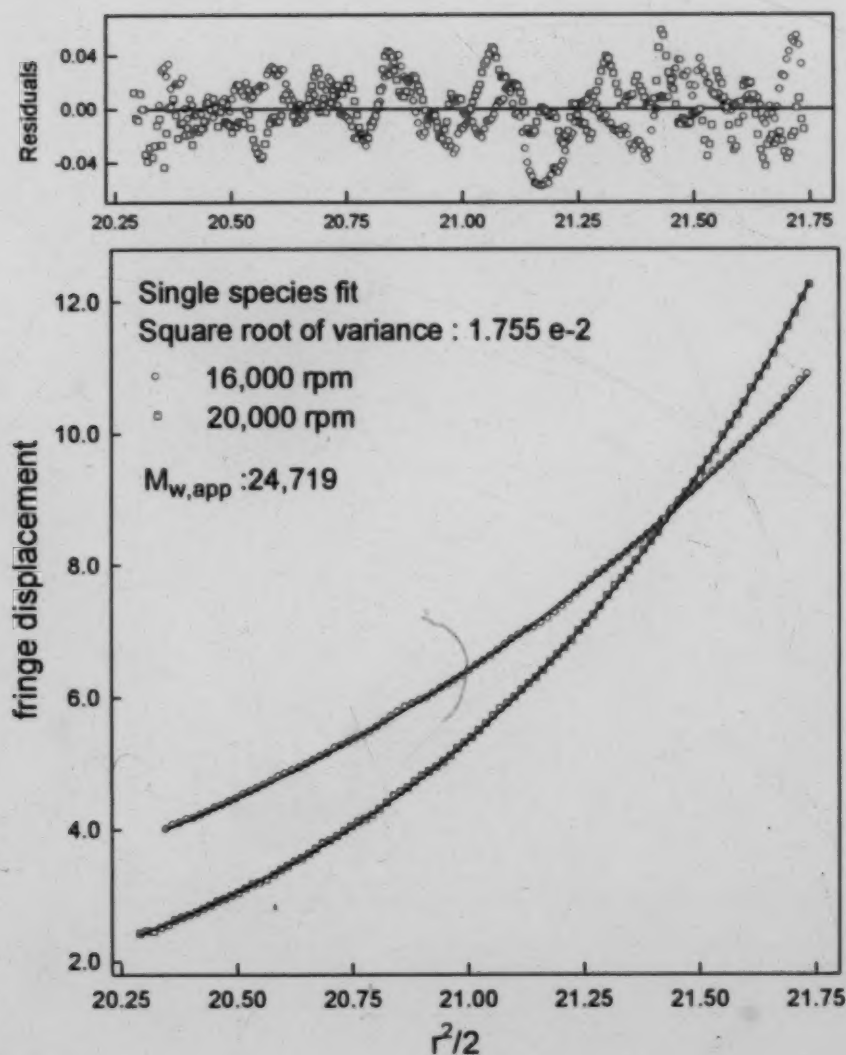


Figure 2.8: Sedimentation Analysis of HAV 3C protease in 100 mM NaCl, 20 mM KH_2PO_4 , 0.5 mM EDTA, 1.0 mM DTT and pH 5.4.

~ 33% as calculated from analytical ultracentrifugation and ^{15}N -T₂ enveloped studies respectively.

Table 2.3. Analysis of the ^{15}N -amide envelope of HAV 3C Protease at four different concentrations and two sets of conditions. τ_c and MW are estimated using equations found in Spyropoulos et al., 1999. ($1.11\tau_c = 1/T_{2\text{NH}}$, $\text{MW} = 2 \times \tau_c$)

	Sample A	Sample B	Sample C	Sample D
^{15}N T ₁ (ms)	~ 1015	~ 1003	~ 944	~ 964
^{15}N T ₂ (ms)	~ 74	~ 59	~ 57	~ 77
τ_c (ns)	~ 12.1	~ 15.3	~ 15.8	~ 11.7
MW (kDa)	~ 24.2	~ 30.6	~ 31.6	~ 23.4

Sample A: 1.40 mM HAV 3C Protease / 100 mM NaCl / 20 mM KH₂PO₄ / 0.5 mM EDTA / 1.0 mM DTT / pH 5.4
Sample B: 1.0 mM HAV 3C Protease / 100 mM NaCl / 20 mM KH₂PO₄ / 0.5 mM EDTA / 1.0 mM DTT / pH 5.4
Sample C: 0.75 mM HAV 3C Protease / 350 mM NaCl / 20 mM KH₂PO₄ / 0.5 mM EDTA / pH 7.0
Sample D: 0.46 mM HAV 3C Protease / 100 mM NaCl / 20 mM KH₂PO₄ / 0.5 mM EDTA / 1.0 mM DTT / pH 5.4

Samples A, B and D in Table 2.3 were studied to determine if increasing protease concentrations would increase ^{15}N -T₂ times which in turn would reflect increased levels of aggregation. Of these three samples, sample B (1.0 mM) yielded results that did not fit with samples A (1.4 mM) and D (0.46 mM). Sample B was prepared from frozen stock and had an estimated ^{15}N -T₂ of ~ 59 ms which is very similar to the ^{15}N T₂ (~ 57 ms) of the protease in the original high salt/high pH conditions. It appears that freezing and thawing prior to sample preparation can lead to problems with aggregation, even when the protein is under optimal low salt, low pH conditions (100 mM NaCl, 20 mM KH₂PO₄, 0.5 mM EDTA, 1.0 mM DTT, and pH 5.4). This observation supports the observations

from the 1D ^1H NMR linewidth studies that protease stock should be kept chilled at 4°C and not frozen at -20°C . Sample D (0.46 mM) and sample A (1.40 mM) had very similar ^{15}N - T_2 times and therefore it appears that increasing the concentration of the protease to maximum of 1.40 mM does not lead to any significant aggregation. It is important to note that the 1.40 mM protease sample had exhibited some precipitation after 48 hours as the sample was cloudy with flakes. Preparing NMR samples at this high concentration may be risky for longer 3D NMR experiments that take three days to complete. In contrast, sample D (0.46 mM) showed no precipitate for several weeks after it was prepared.

Stability Study of the HAV 3C Protease. This stability study attempted to address two key issues; (1) the length of time the protease stock solution would be stable when stored at 4°C and (2) the length of time a concentrated (0.63 mM) sample (stored at 4°C and 25°C) would remain monomeric after preparation. The results of this study are shown in Table 2.4. HAV 3C protease was expressed and purified as described previously and stored in a stock solution (2mg/mL) for a month prior to preparing the two samples used in this study. Spectra collected for the two 0.63 mM samples on days 1 through 9 displayed a half-height linewidth of 16 Hz. Therefore, the concentrated protease sample appears to be stable for nine days regardless of whether it is stored at 4°C or 25°C . The increase in half-height linewidth noted on day 27 (2 Hz at 4°C and 1 Hz at 25°C) could be due to self-association although the extent of aggregation is not known.

Table 2.4. Stability Study of (~ 0.6 mM) HAV 3C protease NMR samples at 4°C and 25°C .

Day #	$\Delta\nu_{1/2}$ (Hz) of Sample at 4°C	$\Delta\nu_{1/2}$ (Hz) of Sample at 25°C
1	16 Hz	16 Hz
4	16 Hz	16 Hz
7	16 Hz	16 Hz
9	16 Hz	16 Hz
27	18 Hz	17 Hz

The results of this study indicate that HAV 3C protease can be stored at 4°C at low concentrations (2 mg/mL) in the elution buffer for at least one month with no evidence of aggregation. This study also shows that the concentrated (0.63 mM) protease samples are stable for at least nine days and that after this time, some degree of 'self-association' is noted. This suggests that concentrated HAV 3C protease samples should be prepared and analyzed by NMR within 10 days of preparation. These samples yield the best results within the first nine days although it may be that well resolved spectra can be recorded up to ~ one month after preparation.

2.4 Conclusion

To summarize, solution conditions have been found where the HAV 3C protease remains monomeric at high (> 0.5 mM) sample concentrations. To identify these conditions a number of solution variables (salt, pH, temp, detergents, addition of ligands) were screened via half-height linewidth measurements. Subsequent studies using two techniques, analytical ultracentrifugation and ^{15}N -T₂ relaxation studies, clearly established that the HAV 3C protease was monomeric at low salt (100 mM NaCl) and low pH (pH 5.4) conditions. Our ^{15}N NMR relaxation studies indicate that NMR samples up to 1.4 mM can remain monomeric although some precipitation typically occurs from more concentrated samples after 48 hours. Finally, stability studies indicate that after the protease has been expressed and purified, it can be stored at ~2-3 mg/mL in the elution buffer at 4°C for up to one month and that concentrated (~0.63 mM) HAV 3C protease samples show no sign of aggregation for at least nine days. This information proved to be vital for the success of the triple resonance experiments that are described later in this thesis.

Chapter 3

^2H , ^{13}C and ^{15}N isotopic labelling of HAV 3C protease

3.1 Introduction

^2H , ^{13}C and ^{15}N isotopic labelling of proteins has become a standard practice in biomolecular NMR spectroscopy. It is primarily used to facilitate the study of larger (>12 kDa) proteins. If a protein can be biosynthesized by a bacterial (or a simple eukaryotic) expression system, ^{15}N and ^{13}C can be readily incorporated by growing the cells in minimal media containing isotopically enriched nitrogen and carbon sources (McIntosh and Dahlquist, 1990; Pardi 1992). Similarly, ^2H can be incorporated by growing cells in $^2\text{H}_2\text{O}$ -based media (Venters et al., 1995). When ^{13}C and ^{15}N are incorporated into proteins, multidimensional heteronuclear NMR experiments can be acquired to alleviate the peak overlap problems often found in 2D ^1H NMR spectra of unlabelled proteins. Concomitantly, these heteronuclear spectra are easier to work with, owing to the large chemical shift dispersion inherent to ^{13}C and ^{15}N nuclei (Clare and Gronenborn, 1991; Kay et al., 1990a; Bax et al., 1989). Incorporation of ^2H into bio-molecules is primarily used to improve the relaxation properties of larger proteins (Gardner and Kay, 1998) leading to better spectral resolution. By using a suite of triple-resonance experiments developed in Dr. Lewis Kay's laboratory for deuterated proteins (Yamazaki et al., 1994a/b), several large proteins including the 29 kDa human carbonic anhydrase II (Venters et al., 1995) and a 37 kDa trp repressor/DNA ternary complex (Yamazaki et al., 1994a/b) have been successfully assigned.

While uniform incorporation of ^2H , ^{13}C and ^{15}N is the predominant strategy for labelling proteins, residue selective ^{15}N labelling is a useful alternative as it can assist in defining which resonances in heteronuclear NMR spectra belong to a

specific amino acid type. This information can be invaluable, as it can be difficult to assign spin systems to residue types on the basis of chemical shift information alone. For example, when using HNCA and HN(CO)CA spectra to obtain backbone sequential chemical shift assignments, ambiguities may arise due to the similarities in $^{13}\text{C}_\alpha$ chemical shifts displayed by certain groups of amino acids. One such group includes histidine, leucine, methionine and glutamine as they all display $^{13}\text{C}_\alpha$ chemical shifts near ~53 ppm (Wishart et al., 1995b). By using ^1H - ^{15}N HSQC spectra collected from residue-specific ^{15}N labelled samples, unambiguous assignment of spin systems to a given residue type is possible, provided that these types of specifically labelled samples can be successfully biosynthesized. This type of labelling strategy has aided in the NMR study of several proteins including α -lytic protease (Bachovchin, 1985), *E. coli* thioredoxin (LeMaster and Richards, 1985), T4 lysozyme (McIntosh et al., 1990) and a M13 coat protein (Henry and Sykes, 1992).

In this study, several biosynthetic strategies were employed to introduce isotopes into the HAV 3C protease. Single (^{15}N), double ($^{13}\text{C}/^{15}\text{N}$), triple ($^2\text{H}/^{13}\text{C}/^{15}\text{N}$) and residue-specific (^{15}N) labelled protease samples were all successfully produced. This was accomplished by growing the *E. coli* cells containing the HAV 3C protease expression system in $^1\text{H}_2\text{O}$ - or $^2\text{H}_2\text{O}$ -based minimal media supplemented with $^{15}\text{NH}_4\text{Cl}$, $^{13}\text{C}_6$ -glucose and/or amino acid mixtures containing ^{15}N -amino acids. These strategies produced protease samples with high levels (>95%) of ^{13}C and ^{15}N and moderate levels (~70 %) of ^2H incorporation. In addition, a simple residue-specific labelling strategy with a prototrophic strain of *E. coli*, produced "good" results with HAV 3C protease labelled specifically at alanine, leucine, isoleucine, lysine, and valine.

3.2 Materials and Methods

Isotopes and Reagents. $^{15}\text{NH}_4\text{Cl}$ (98%), ^{13}C -C₆-Glucose (>98%), and sterile filtered $^2\text{H}_2\text{O}$ (>97%) were purchased from Martek Biosciences Corporation (Columbia, MD). ^{15}N -Alanine, ^{15}N -Glycine, ^{15}N -Isoleucine, ^{15}N -Leucine, ^{15}N -Lysine, ^{15}N -Phenylalanine, and ^{15}N -Valine (all 95-99%) were purchased from Cambridge Isotope Laboratories, Inc (Andover, MA). Unlabelled amino acids were obtained from Sigma Chemical Co. (St. Louis, MO). Reagents used for the expression and purification of HAV 3C protease are listed in Table 3.1.

Table 3.1. List of reagents used for the expression and purification of HAV 3C protease.

Reagent	Ingredient	Quantity
M9 Minimal Media Add ingredients to sufficient double distilled (dd) H ₂ O to make 1 L.	NaHPO ₄	6.80 gm
	NaCl	0.50 gm
	NH ₄ Cl	1.00 gm
	Glucose	4.00 gm
	MgSO ₄ (1.0M)	2.00 mL
	CaCl ₂ •2H ₂ O (1.0M)	0.10 mL
	Thiamine (1% w/v)	1.00 mL
	Ampicillin (75mg/mL)	2.00 mL
	FeSO ₄ (1 mM)	2.00 mL
Cell Resuspension Buffer Add ingredients to sufficient dd H ₂ O to make 100 mL.	Tris-HCl (1.0M / pH 8.5)	5.00 mL
	EDTA (0.5 M / pH 8.0)	0.50 mL
	DTT (1.0M / pH 5.2)	0.20 mL
Phosphate Buffer Add ingredients to sufficient dd H ₂ O to make 1 L. Correct solution pH to 5.4 with 1.0 M KOH or 1.0 M HCl.	KH ₂ PO ₄	2.72 gm
	0.5 M EDTA (pH 8.0)	1.00 mL
	1.0 M DTT (pH 5.2)	2.00 mL

HAV 3C protease expression and purification. All HAV 3C protease samples were isolated from a prototrophic *E. coli* strain MM294 containing the overexpression plasmid pHAV-3CEX (Malcolm et al., 1993) with a C24S mutation (to facilitate protein stability). The four types of isotopically labelled HAV 3C protease have been prepared by growing the *E. coli* expression system in four different minimal media preparations as described below. The expression and purification of HAV 3C protease was performed as outlined in Chapter 2. Expected yields and parameters used to obtain each type of isotopically labelled HAV 3C protease sample are outlined in Table 3.2.

Table 3.2. Comparison of strategies used in the isolation of HAV 3C protease. T1 is the approximate time used for the overnight culture growth. T2 is the approximate time between inoculation of the "expression" media and induction with IPTG. T3 is the "induction period" that cells were allowed to grow prior to harvesting. M9* is minimal media (1 L) prepared with 1.0 gm of ^{13}C -C₆-glucose and 1.0 gm $^{15}\text{NH}_4\text{Cl}$. M9** is minimal media (1 L) prepared with an amino acid mixture substituted for NH_4Cl . M9*** is minimal media (300 mL) prepared with 90 % $^2\text{H}_2\text{O}$:10% $^1\text{H}_2\text{O}$, 0.6 gm of ^{13}C -C₆-glucose and 0.3 gm of $^{15}\text{NH}_4\text{Cl}$.

Isotopic Label	Overnight media	Expression media	T1 (hrs)	T2 (hrs)	T3 (hrs)	Yield of HAV 3C protease (mg/L)
^{15}N	TB	M9	12-14	2-4	8	~ 20
$^{13}\text{C}/^{15}\text{N}$	TB	M9*	12-14	2-4	8	~ 8
^{15}N residue specific	TB	M9**	12-14	2-4	8	~ 40
Unlabelled	M9	M9	12-14	2-4	8	~ 80
$^2\text{H}/^{13}\text{C}/^{15}\text{N}$	M9***	M9***	48	6	12	~ 25

^{15}N uniformly labelled HAV 3C protease. To uniformly incorporate ^{15}N into HAV 3C protease, 1 L of M9 minimal media was prepared by substituting 1 gm of $^{15}\text{NH}_4\text{Cl}$ for 1 gm of unlabelled NH_4Cl . *E. coli* cells (containing the overexpression plasmid pHAV-3CEX) were grown overnight in a 25 mL rich

media (TB) starter culture. If growth occurred, 10 mL of this starter culture was used to inoculate the 1 L batch of $^{15}\text{NH}_4\text{Cl}$ -minimal media. The *E. coli* cells utilized $^{15}\text{NH}_4\text{Cl}$ similarly to unlabelled NH_4Cl therefore no adaptation period or change in protocol was necessary. Therefore, isolation and purification of the uniformly ^{15}N labelled HAV 3C protease was done as outlined in Chapter 2.

$^{13}\text{C}/^{15}\text{N}$ uniformly labelled HAV 3C protease. In order to limit the expense of ^{13}C - C_6 -glucose used to prepare a doubly labelled $^{13}\text{C}/^{15}\text{N}$ HAV 3C protease, several trials were performed to determine the minimal amount of glucose (per 1 L batch) that the *E. coli* cells could successfully grow on. These studies indicated that a sufficient quantity of HAV 3C protease (for a NMR sample) could be obtained by growing cells in a 1 L batch of minimal media having 1.0 gm glucose (data not shown). With this information, 1 L of M9 minimal media was prepared by substituting 1.0 gm of $^{15}\text{NH}_4\text{Cl}$ for 1.0 gm of unlabelled NH_4Cl and 1.0 gm of ^{13}C - C_6 -glucose for 4 gm of glucose. As before, *E. coli* cells were grown overnight in a 25 mL rich media starter culture. If growth occurred, 10 mL of this starter culture was used to inoculate the 1 L batch of minimal media. The *E. coli* cells utilized ^{13}C - C_6 -glucose similarly to unlabelled glucose therefore no adaptation period or change in protocol was necessary. Therefore, isolation of the $^{13}\text{C}/^{15}\text{N}$ labelled HAV 3C protease was done as outlined in Chapter 2.

$^2\text{H}/^{13}\text{C}/^{15}\text{N}$ uniformly labelled HAV 3C protease. Triple labelling ($^2\text{H}/^{13}\text{C}/^{15}\text{N}$) HAV 3C protease with a high level of deuterium required an initial adaptation period as outlined in Figure 3.1. This involved growing our *E. coli* expression system in minimal media prepared with increasing concentrations of sterile filtered $^2\text{H}_2\text{O}$ (>97%). Two mL quantities of minimal media, composed of 50, 80, and 85 percent $^2\text{H}_2\text{O}$, were prepared and filtered through a 0.22 μm Millex GV filter (Millipore) to ensure sterilization. A single colony was selected from a LB/amp plate and transferred to minimal media containing 50% $^2\text{H}_2\text{O}$:50% $^1\text{H}_2\text{O}$. This mixture was grown at 37°C for 20 hours at which time the OD_{600} reached 1.6. It was then stored overnight at 4°C. The next day, 20 μL of this mixture was added

to two additional (2.0 mL) media preparations having 80 and 85 percent $^2\text{H}_2\text{O}$. The 80 percent $^2\text{H}_2\text{O}$ minimal media was used to ensure the expression system had not lost the HAV 3C protease vector. This culture was grown at 37°C and when the OD_{600} of this culture reached 0.6, IPTG was added to a final concentration of 2 mM to induce the overexpression of HAV 3C protease. Immediately after induction, the temperature was lowered to 30°C and the cells were allowed to grow for a period of 8 hours. The cells were collected and lysed via the freeze-thaw method described in Chapter 2. The presence of HAV 3C protease was confirmed by SDS-Page gel electrophoresis. The 85 percent $^2\text{H}_2\text{O}$ culture was grown for 20 hours at 37°C to an OD_{600} of 1.8. The $^2\text{H}_2\text{O}$ -adapted MM294 cells from this culture were used to prepare several -80°C freezer stocks by mixing 1 mL of the culture with 1 mL of sterile 30 percent glycerol.

To triple label ($^2\text{H}/^{13}\text{C}/^{15}\text{N}$) the protease, 300 mL of 90% $^2\text{H}_2\text{O}$:10% $^1\text{H}_2\text{O}$ based minimal media was prepared using only 0.6 gm ^{13}C - C_6 -glucose and 0.3 gm $^{15}\text{NH}_4\text{Cl}$. Cells taken from the " $^2\text{H}_2\text{O}$ adapted" -80°C freezer stock were transferred directly into 30 mL of this deuterated minimal media. Due to the slow growth in deuterated minimal media (Table 3.2, p.71), the 12-14 hour period for previous overnight cultures was extended to 48 hours. After this period, the entire volume (30 mL) of overnight culture was transferred into the remaining 270 mL of prepared minimal media. In approximately six hours the OD_{600} reached 0.77 and IPTG was added to a final concentration of 2 mM to induce the culture. The temperature was lowered to 30°C and growth was allowed to continue for an extended induction period of 12 hours. After this stage, the triply labelled HAV 3C protease was isolated and purified as described previously in Chapter 2.

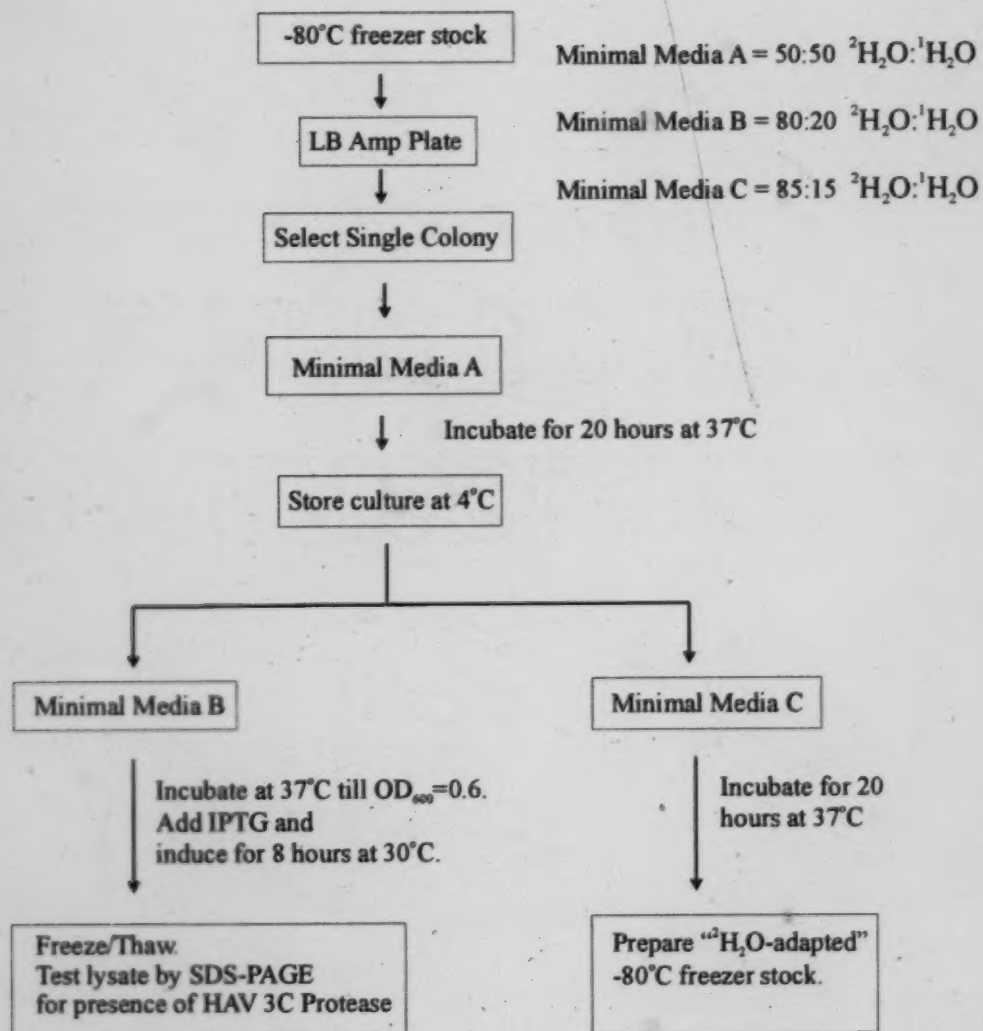


Figure 3.1. Protocol to adapt bacteria to $^2\text{H}_2\text{O}:\text{}^1\text{H}_2\text{O}$ based minimal media.

Residue-specific ^{15}N Labelling of HAV 3C Protease. HAV 3C protease was labelled at specific residues (alanine, glycine, leucine, lysine, isoleucine, phenylalanine, and valine) by growing *E. coli* cells (containing the overexpression plasmid pHAV-3CEX) in a 1 L batch of minimal media supplemented with 100 mL of an amino acid mixture instead of NH_4Cl (Muchmore et al., 1989). This amino acid mixture consisted of 19 unlabelled L-amino acids and one ^{15}N -L-amino acid that had to be specially prepared as outlined in Table 3.3. 10 mL of each of the 10 amino acid solutions (100mg/10 mL) was combined and filter sterilized through a 0.22 μm Millex GV filter (Millipore) to ensure sterilization. As before, *E. coli* cells were grown overnight in a 25 mL rich media starter culture. If growth occurred, 10 mL of this starter culture was used to inoculate this 1 L batch of minimal media. The *E. coli* cells grew efficiently on this "amino acid" minimal media producing greater quantities (Table 3.2) of HAV 3C protease than seen before although no adaptation period or change in protocol was necessary. Therefore, isolation of the residue-specific ^{15}N labelled HAV 3C protease followed identically the protocol outlined in Chapter 2.

2D ^{15}N - ^1H HSQC studies of uniform and residue-specific labelled HAV 3C protease. These experiments were performed by Dr. Carolyn Slupsky and Dr. Brian Sykes. NMR samples (~1.0 mM) of all residue specific and uniform labelled HAV 3C protease were prepared in 90% $^1\text{H}_2\text{O}$:10% $^2\text{H}_2\text{O}$, 350 mM NaCl, 20 mM KH_2PO_4 , 0.5 mM EDTA at pH 6.8 except for the uniform and ^{15}N l-isoleucine sample that were prepared in 90% $^1\text{H}_2\text{O}$:10% $^2\text{H}_2\text{O}$, 100 mM NaCl, 20 mM KH_2PO_4 , 0.5 mM EDTA, 1.0 mM DTT at pH 5.4. ^{15}N - ^1H HSQC experiments (Kay et al., 1992; Zhang et al., 1994) were carried out at 25°C on a Varian Unity 600 spectrometer equipped with three channels and a 5 mm pulse-field gradient triple resonance probe. The spectra were acquired with 64 to 128 complex points in the F1 (^{15}N) dimension and 512 complex points in the F2 (^1H) dimension. Depending on the concentration and quality of the sample 96-224 transients were collected. The spectral widths were typically 2000 Hz in F1 (^{15}N) and 8000 Hz in the F2 (^1H) dimension. Samples were directly (^1H) and indirectly (^{15}N) referenced

to DSS as described by Wishart et al. (1995). Spectra were processed and peak-picked on a Sun Ultra 1 workstation, using the NMRPipe (Delaglio et al., 1995) and PIPP (Garrett et al., 1991) software packages.

Table 3.3. Preparation of 1% (w/v) amino acid stock solutions. 500 mg of each amino acid in each amino acid group was dissolved in sufficient dd H₂O to make 50 mL. Solutions were filtered through a 0.22 µm Millex GV filter (Millipore) to ensure sterilization. Store all at 4°C except for Cys-Asn-Gln which is stored at room temperature.

Amino Acid Group	Special Requirements
Trp	Add sufficient drops of 1.0 M HCl to dissolve.
Tyr	Add sufficient drops of 1.0 M KOH to dissolve.
Asp-Glu	Add sufficient drops of 1.0 M KOH to dissolve.
Cys-Asn-Gln	
Leu-Ile-Val	
Met	
Ala-Thr-Phe-Pro-Ser	
Arg-Lys	
Gly	
His	

Mass Spectral Analysis. Four 100 µL aliquots of HAV 3C protease (non-labelled, ¹⁵N-labelled, ¹³C/¹⁵N-labelled and ²H/¹³C/¹⁵N-labelled) were examined via mass spectral analysis. These aliquots (0.5 - 2.0 mg/mL HAV 3C protease) were in the post-column elution buffer which was either 350 mM NaCl, 20 mM KH₂PO₄, 0.5 mM EDTA at pH 6.8 or 350 mM NaCl, 20 mM KH₂PO₄, 0.5 mM EDTA, 1.0 mM DTT at pH 5.4. The 100 µL aliquot was taken from the purest post-column fraction (containing HAV 3C protease) as judged by SDS-PAGE. The four protease samples were submitted to PENCE "as-is" to determine the experimental mass of each sample by electrospray mass spectrometry using a Fisons VG Quattro LCMS. The obtained experimental masses are reported in Table 3.4.

3.3 Results and Discussion

Adapting bacteria to chemically defined media. Initially, the protocol for the expression and purification of HAV 3C protease involved growing our bacterial expression system in a rich (TB) media overnight. 10 mL of this culture was used to inoculate 300 mL to 1 L of minimal media ("expression" media) prepared with ^{13}C or ^{15}N enriched components. This protocol usually led to the expression of HAV 3C protease in yields of ~ 15-25 mg/L.

When attempting to label the protein with $^2\text{H}/^{13}\text{C}/^{15}\text{N}$, concerns arose over the use of the rich media as an "overnight" culture and the possible carry over of protonated nutrients to the deuterated media. To eliminate this problem, we tested and later confirmed that minimal media could be used in the initial starter culture. Interestingly, 12 months prior to this, many attempts to grow the same cells in minimal media proved unsuccessful. One explanation for our later success could be that at some point in the study, the MM294 cells harboring the pHAV-3CEX vector had mutated allowing them to grow better in minimal media. Conversely, the latter efforts differed from the earlier efforts in that a much longer growth period was used to allow cellular adaptation to the minimal media. Regardless, the use of minimal media in the starter culture proved to be very fruitful as it led to significantly higher yields (~ 75-100 mg/l) of HAV 3C protease as compared to those yields seen before (Table 3.2, p.71). The improved yields of HAV 3C protease is likely the result of adapting bacteria to minimal media prior to using them to inoculate the 1.0 L minimal ("expression") media. This agrees with Mossakowska and Smith (1997) who report that swapping from rich complex media, in which all the *E. coli* requirements are provided, to minimal media in which *E. coli* have to synthesize key intermediary metabolites, usually results in much poorer and slower growth.

Table 3.4. ^2H , ^{13}C and ^{15}N incorporation levels in singly, doubly and triply labelled HAV 3C protease as determined by electrospray mass spectrometry.

Isotopic Label	Theoretical Mass	Observed Mass	% isotope enrichment
Unlabelled	23 876.59	23 923.92 $\pm$ 13.78	--
^{15}N	24 165.36	24 149.39 $\pm$ 12.35	$\sim$ 95.5 (^{15}N)
$^{13}\text{C}/^{15}\text{N}$	25 240.77	25 184.76 $\pm$ 15.82	$\sim$ 95.5 (^{15}N) $\sim$ 96.3 (^{13}C)
$^2\text{H}/^{13}\text{C}/^{15}\text{N}$	26 621.00	26 184.01 $\pm$ 19.66	$\sim$ 95.5 (^{15}N) $\sim$ 96.3 (^{13}C) $\sim$ 72.4 (^2H)

Isotopic Incorporation. Electrospray mass spectrometry was used to determine the mass for the unlabelled and three uniformly labelled types of HAV 3C protease. These results are reported in Table 3.4. The mass of the unlabelled protease was measured to be 23,923.92 $\pm$ 13.78 which is in good agreement ($\pm$ 1 % or $\pm$ 23.8 Da) with the theoretical mass of 23,876.59 based on the amino acid sequence and the natural abundance of all isotopes. The protease contains 287 nitrogen nuclei therefore if all ^{14}N nuclei were replaced by ^{15}N nuclei the mass should increase by 288.77 mass units. The measured mass of 24,149.39 $\pm$ 12.35 for the ^{15}N labelled protease is 15.97 mass units less than the theoretical mass therefore ^{15}N incorporation was estimated to be (272.80/288.77) 95.5 percent. The protease contains 1069 carbon nuclei therefore replacing all carbons with ^{13}C should increase the mass by 1075.41 mass units. The measured mass of 25,184 $\pm$ 15.82 for the $^{13}\text{C}/^{15}\text{N}$ labelled protease is 56.01 mass units less than the expected mass. Assuming that 15.97 mass units could be attributed to ^{15}N deficiency, the incorporation of ^{13}C was said to be (1035.37/1075.41) 96.3 percent. The protease has $\sim$ 1707 protons (which varies with pH) therefore if all except for the 335 attached to nitrogen (backbone and sidechain) were exchanged with ^2H , the protease would increase in mass by 1380.23 mass units. An observed mass of 26,184.01 $\pm$ 19.66 was measured for the triply labelled protease which is 437.00 mass units less than the theoretical. Assuming that 56.01 could be attributed to ^{15}N

and ^{13}C deficiencies, the level of isotopic incorporation for ^2H was said to be (999.24/1380.23) 72.39 percent.

The less-than-perfect (96 %) labelling for the uniformly ^{15}N and $^{13}\text{C}/^{15}\text{N}$ labelled proteases could be attributed to the contamination arising from the 10 mL of unlabelled TB used to inoculate the 1L minimal media "expression" media. However, this level of incorporation was deemed satisfactory since any level >90% is the normal requirement for satisfactory heteronuclear experiments (Mossakowska and Smith, 1997). For the triply labelled sample, deuterium incorporation was estimated to be approximately 70 % which, initially seemed quite low considering the M9 media was composed of 90 % $^2\text{H}_2\text{O}$. Usually the level of ^2H incorporation reflects the percentage of $^2\text{H}_2\text{O}$ used to prepare the minimal media (Gardner and Kay, 1998). Therefore, using 90 % $^2\text{H}_2\text{O}$ minimal media we should have expected approximately 90 % ^2H incorporation. It is thought that several factors lead to this less-than-expected distribution of deuterium throughout the protease. One factor could be that protonated glucose instead of deuterated glucose was used. This is of importance because during biosynthesis, protons from these molecules can be efficiently retained at the side chains of aromatic amino acids and other specific sites within several amino acids (Gardner and Kay, 1998). Additionally, metabolic and biosynthetic enzymes may have preferentially used protons over deuterons, since both were in the medium. This preference is thought to be due to kinetic and thermodynamic isotope effects (Galimov, 1985). Regardless, this level of deuteration was deemed to be satisfactory since other proteins, such as a 37 kDa trp repressor/DNA ternary complex (Yamazaki et al, 1994b), have been successfully studied at ^2H incorporation levels of approximately 70 %.

^{15}N Labelling at Specific Amino Acids in the HAV 3C Protease. Amino acid biosynthesis in bacteria is regulated by enzymatic activity and at the level of gene expression (Neidhardt, 1987). Endogenous synthesis of unlabelled amino acids and transaminase-catalysed nitrogen exchange often can be minimized by

supplying the host bacteria with sufficient quantities of various amino acids in the "expression" medium (McIntosh and Dahlquist, 1990). Therefore the simplest approach, and the strategy we used in this study was to use a defined "expression" medium (minimal media minus NH_4Cl) containing one ^{15}N labelled amino acid mixed with 19 unlabelled amino acids. The extent to which a ^{15}N labelled amino acid will be incorporated depends on many factors including the amino acid in question and the growth characteristics of the bacterial host.

The ^1H - ^{15}N HSQCs recorded for the ^{15}N specifically labelled alanine, glycine, isoleucine, leucine, lysine, phenylalanine and valine HAV 3C protease samples are shown in Figures 3.2 (p.81) and 3.3 (p.82). Inspection of this group of ^1H - ^{15}N HSQC spectra will allow one to discern two characteristics that are generally conserved. First, all seven spectra show some degree of ^{15}N being incorporated into the side chain nitrogen nuclei of arginine, lysine, glutamine, and asparagine. This is evident as the "hourglass-shaped" peaks between 6.4 and 7.8 ppm in the ^1H dimension and 110 and 114 ppm in the ^{15}N dimension. Second, in addition to the strong peaks, which are assumed to be the intended amino acid, there are usually a few extra less intense peaks found in each spectrum. One of these extra peaks ($^{15}\text{N} = 130.20$ ppm, $^1\text{H} = 8.01$ ppm) corresponds to the C-terminal residue E217. Likewise, there are three other peaks (120.80/7.72, 123.95/8.40, 123.99/8.32) found in most of the spectra corresponding to the other C-terminal residues K214, K215, and I216.

This non-specific ^{15}N incorporation, otherwise known as "scrambling", can be caused by enzymes that convert the exogenously supplied ^{15}N labelled amino acid into a metabolic precursor or it can be caused by transaminases which catalyze nitrogen exchange into undesired sites of other amino acids (Muchmore et al., 1989). It is likely that all of the specific labels experienced "scrambling" to some degree since an auxotrophic strain of bacteria was not used to prepare the specifically labelled samples. The reason that the C-terminal resonances are

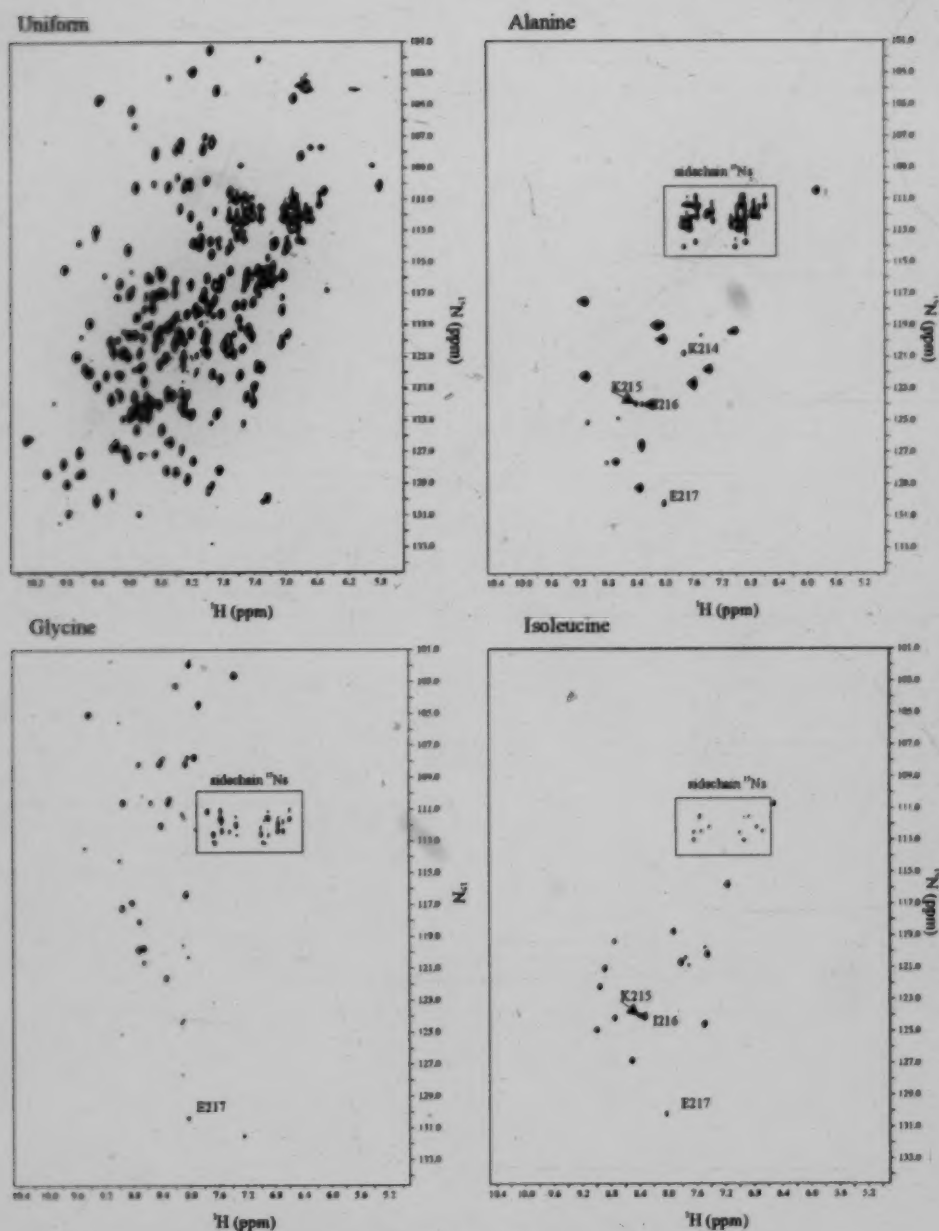


Figure 3.2. ^1H - ^{15}N HSQC spectra of (~ 1 mM) uniform ^{15}N and ^{15}N -isoleucine labelled HAV 3C protease in 'low salt-low pH' conditions and ^{15}N -alanine and ^{15}N -glycine labelled HAV 3C protease in 'high salt-high pH' conditions collected at 25°C on a Varian Unity 600 spectrometer.

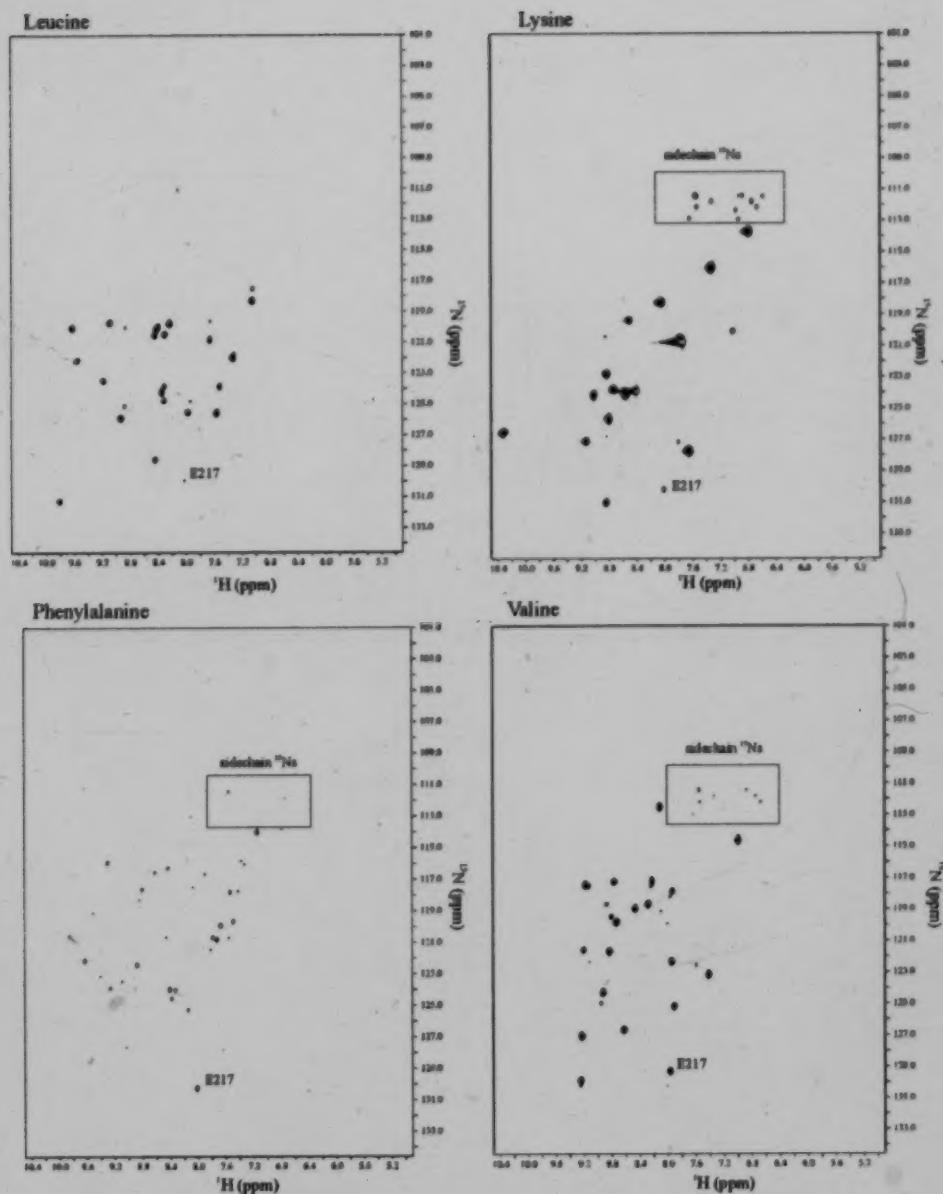


Figure 3.3. ^1H - ^{15}N HSQC spectra of (~ 1 mM) ^{15}N -leucine, ^{15}N -lysine, ^{15}N -phenylalanine and ^{15}N -valine labelled HAV 3C protease in 'high salt-high pH' solution conditions collected at 25°C on a Varian Unity 600 MHz spectrometer.

visible is likely due to their very high mobility. Because of this mobility, they display very sharp intense peaks that exaggerate their level of ^{15}N incorporation.

The relative success of this simple specific labelling strategy (without the use of auxotrophic strains of *E. coli*) is reported in Table 3.5. Five of the seven trials gave very good results (showing greater than 80% of the expected peaks). In particular, the spectra collected for alanine, isoleucine, leucine, lysine, and valine labelled HAV 3C protease were generally good. Of these, the isoleucine and leucine samples gave the best results (although leucine must have scrambled preferentially into one other residue in the sequence) showing 100% of the expected peaks. The reason that these two samples displayed all of the expected peaks is likely due to the fact that they were prepared at a lower pH (5.4) which led to a slower rate of hydrogen exchange and improved amide ($^1\text{H}/^{15}\text{N}$) signals. It is possible that if the other samples had been prepared at this lower pH, they too would have displayed all (or at least more) of the expected peaks. Attempts to label HAV 3C protease at glycine and phenylalanine were the least useful as it is quite clear that scrambling occurred for both samples. However, of these two, the glycine spectrum was of some utility as 15 of the 22 expected peaks could still be discerned.

Table 3.5. Success of residue-specific ^{15}N labelling studies.

Amino Acid type	Total number in sequence	Number of intense peaks in $^1\text{H}-^{15}\text{N}$ HSQC spectrum
Alanine	12	11
Glycine	22	"scrambled"
Isoleucine	13	13
Leucine	19	20
Lysine	17	15
Phenylalanine	8	"scrambled"
Valine	23	18

3.4 Conclusion

To summarize, nearly all of our isotopic labelling attempts proved to be very successful. For the three uniformly labelled protease samples ($^2\text{H}/^{13}\text{C}/^{15}\text{N}$, $^{13}\text{C}/^{15}\text{N}$, and ^{15}N), incorporation levels of >95% were noted for both ^{13}C and ^{15}N and ~ 70 % for ^2H . It was determined that careful adaptation of the bacterial expression system to chemically defined media played a major role in optimizing our yields of the labelled HAV 3C protease. This was demonstrated in the higher yields (75-100 mg/L versus 20 mg/L) of protease when using minimal media instead of rich media in the "overnight" culture. Cellular adaptation was also important in producing HAV 3C protease in deuterated media. This was demonstrated in the success of the step-wise approach to progressively higher percentages of $^2\text{H}_2\text{O}$ content which we used to select " $^2\text{H}_2\text{O}$ -adapted" bacteria in the production of the triple ($^2\text{H}/^{13}\text{C}/^{15}\text{N}$) labelled protease. Finally, the ^{15}N residue-specific labelling schemes, even without the use auxotrophic strains of *E. coli*, were successful in five of the seven attempted residue-specific trials.

Chapter 4

¹H, ¹³C and ¹⁵N backbone assignments of HAV 3C protease

4.1 Introduction

Preclinical drug discovery generally involves two processes: 1) lead generation and 2) lead optimization. In the lead generation process, candidate molecules are identified that appear to bind to a protein target and inhibit the *in vitro* activity of the protein. In the lead optimization process, potential drug leads are optimized with respect to detailed structural information, *in vitro* potency and other important parameters such as bioavailability, pharmacokinetic or toxicological properties. In recent years, NMR spectroscopy has been applied to both lead generation and lead optimization (Moore, 1999; Roberts, 1999; Fesik, 1993; Hajduk et al., 1997; Shuker et al., 1996). In the area of lead generation, NMR has been used to detect weak binding of small molecules to a number of protein drug targets. Subsequently, this binding information can be used to design much tighter-binding inhibitors. The method, known as "SAR by NMR" (Shuker et al., 1996), screens a large library (10^3 - 10^4) of small molecules against an ¹⁵N labelled protein target by using information derived from 2D ¹H-¹⁵N HSQC spectra. Using this technique, protein-ligand binding can be discerned by noting perturbations in the target protein's amide chemical shifts. In the area of lead optimization, NMR has been used to characterize the detailed structure and dynamics of both the protein target and the small molecule leads. Such studies typically involve the solution structure determination of either the protein, the protein-ligand complex or the ligand structure using either isotope editing/filtering or transferred NOE techniques (Roberts, 1999).

For NMR spectroscopy to be used in either of the processes described above, it is critical that complete backbone ¹H, ¹³C and ¹⁵N chemical shift

assignments be available for the protein target of interest. This is because the sequential assignments serve as the basis of any strategy employed to solve protein 3D structure via NMR spectroscopy (Clore and Gronenborn, 1994). Additionally, having the backbone amide ($^1\text{H}_\text{N}$, ^{15}N) chemical shifts assigned in the ^1H - ^{15}N HSQC spectra is of utmost importance for "SAR by NMR" because knowing which residue(s) experience amide chemical shift perturbations indicate which site(s) on the protein are involved in ligand binding (Shuker et al., 1996).

This chapter describes the NMR experiments that were performed to obtain nearly complete ^1H , ^{13}C and ^{15}N backbone chemical shift assignments for HAV 3C protease. Early efforts using ^{15}N -edited and triple resonance experiments (^1H - ^{15}N TOCSY-HSQC, ^1H - ^{15}N NOESY-HSQC, HNCACB and CBCA(CO)NNH) using single (^{15}N) and double ($^{13}\text{C}/^{15}\text{N}$) labelled HAV 3C protease were thwarted by problems with macromolecular aggregation in 'high salt/high pH' solution conditions. The efficient relaxation rates typical of aggregated proteins led to spectra that were devoid of much of the expected chemical shift information. However, identifying solution conditions that minimized aggregation (low salt/low pH) and preparing a triply ($^2\text{H}/^{13}\text{C}/^{15}\text{N}$) labelled HAV 3C protease sample (to combat efficient relaxation rates) solved the "poor spectral quality" problem. With this triple labelled sample, we collected three "deuterium-modified" triple resonance experiments that yielded spectra displaying nearly all of the expected chemical shifts. As a result, these final CT-HNCA, CT-HN(CO)CA and HN(CA)CB spectra were used in conjunction with previously collected residue-specific HSQC data to complete about 85 percent of the sequential backbone assignments.

The assignments were found to be in good agreement with a set of predicted chemical shifts calculated from SHIFTX (Neal, Nip and Wishart, unpublished) and TOTAL (Williamson et al., 1992,1995). These programs use semi-empirical methods (using ring-current shifts, shifts arising from magnetic anisotropies of bonds, and shifts arising from the polarizing effect of polar atoms

on the C H bond) to calculate ^1H , ^{13}C and ^{15}N chemical shifts for HAV 3C protease based on the 3D coordinates from its X-ray crystal structure (Bergmann et al., 1997). Finally, the backbone chemical shifts that have been assigned were analyzed by the CSI program (Wishart et al., 1994) and preliminary calculations show that the locations of 2 α -helices and 11 β -strands are in good agreement with X-ray crystallographic data.

4.2 Heteronuclear NMR experiments used to obtain sequential backbone assignments for HAV 3C protease

Due to the complexity and number of NMR experiments performed in these studies, a concise description of each experiment follows. These experiments include: 1) ^1H - ^{15}N HSQC (Kay et al., 1992; Zhang et al., 1994), 2) ^1H - ^{15}N TOCSY-HSQC (Zhang et al., 1994), 3) ^1H - ^{15}N NOESY-HSQC (Zhang et al., 1994), 4) (constant-time) CT-HNCO (Muhandiram and Kay, 1994), 5) deuterium-modified CT-HNCA (Yamazaki et al., 1994a), 6) deuterium-modified CT-HN(CO)CA (Yamazaki et al., 1994b), and 7) deuterium-modified HN(CA)CB (Yamazaki et al., 1994b) experiments.

2D ^1H - ^{15}N HSQC. 2D ^{15}N - ^1H HSQC experiments (Kay et al., 1992; Zhang et al., 1994) were used to correlate the chemical shift of each ^{15}N nucleus and its covalently bonded proton ($^1\text{H}_\text{N}$). In the ^{15}N - ^1H HSQC experiment, depicted in Figure 4.1 (p.93), magnetization is initially generated on $^1\text{H}_\text{N}$ [Step 1]. An INEPT pulse sequence transfers magnetization to the attached ^{15}N via $^1J_{\text{NH}}$ (~90 Hz) coupling [Step 2] and all ^{15}N chemical shifts are recorded over the acquisition time t_1 [Step 3]. A reverse INEPT pulse sequence transfers magnetization back to the attached $^1\text{H}_\text{N}$ [Step 4] and all $^1\text{H}_\text{N}$ chemical shifts are collected over the second acquisition time t_2 [Step 5]. Fourier transformation of the resultant FID yields a frequency domain map showing all $^1\text{H}_\text{N}$ and ^{15}N correlations.

3D ^1H - ^{15}N TOCSY-HSQC. The 3D ^1H - ^{15}N TOCSY-HSQC experiment (Zhang et al., 1994) is illustrated in Figure 4.2 (p.94). Presaturation of the water resonance was avoided by using 'water-selective pulses' and gradients and by ensuring that water magnetization is placed along the +z-axis prior to acquisition (Grzesiek and Bax, 1993; Kay et al., 1994). Compared to ^1H - ^{15}N HSQC-TOCSY experiments that utilize water presaturation, the advantage of the TOCSY-HSQC is that it avoids saturation of the $^1\text{H}_\alpha$ spins, allowing observation of important $^1\text{H}_\alpha$ cross peaks. Initially this experiment is a homonuclear TOCSY experiment with ^{15}N spin decoupling during the t_1 period [Step 1]. Isotropic mixing transfers magnetization between ^1H spins within each spin system and ^1H chemical shifts are recorded over the acquisition time t_1 . An INEPT pulse sequence then transfers magnetization to ^{15}N via $^1J_{\text{NH}}$ (~90 Hz) coupling [Step 2] and all ^{15}N chemical shifts are recorded over the acquisition time t_2 [Step 3]. Finally, a reverse INEPT pulse sequence is used to transfer magnetization back to $^1\text{H}_\text{N}$ [Step 4] and all $^1\text{H}_\text{N}$ chemical shifts are collected over the final acquisition time t_3 [Step 5]. Fourier transformation of the resultant FID yields a frequency domain map showing all $^1\text{H}_\text{N}$, ^{15}N and ^1H resonances within each spin system.

3D ^1H - ^{15}N NOESY-HSQC. The 3D ^1H - ^{15}N NOESY-HSQC experiment (Zhang et al., 1994) is illustrated in Figure 4.3 (p.95). Presaturation of the water resonance was avoided by using 'water-selective pulses' and gradients and by ensuring that water magnetization is placed along the +z-axis prior to acquisition (Grzesiek and Bax, 1993; Kay et al., 1994). The advantage discussed above for the TOCSY-HSQC applies to the NOESY-HSQC as well. Initially the sequence is a homonuclear NOESY experiment with ^{15}N spin decoupling during the t_1 period [Step 1]. Magnetization transfer occurs via dipolar mediated cross-relaxation for a mixing period τ_M . All NOEs (through space magnetization transfers to ^1H nuclei within 5 Å of each other) are recorded over the acquisition period t_1 . Although all protons are frequency-labelled during t_1 , only magnetization from $^1\text{H}_\text{N}$ is transferred to the attached ^{15}N by an INEPT pulse sequence via the

$^1J_{\text{NH}}$ (~90 Hz) coupling [Step 2]. All ^{15}N chemical shifts are recorded over the acquisition time t_2 [Step 3]. A reverse INEPT pulse sequence is used to transfer magnetization back to the attached $^1\text{H}_\text{N}$ [Steps 4] and all $^1\text{H}_\text{N}$ chemical shift are collected over the final acquisition time t_3 [Step 5]. Fourier transformation of the resultant FID yields a frequency domain map showing all ^{15}N and $^1\text{H}_\text{N}$ correlations as well as ^1H correlations that cross-relaxed with each $^1\text{H}_\text{N}$.

3D CT-HNCO Experiment. The 3D CT-HNCO experiment (Muhandiram and Kay, 1994) was used to correlate the $^1\text{H}_\text{N}$ and ^{15}N chemical shifts of residue 'i' with the carbonyl carbon (^{13}CO) chemical shift of the preceding residue 'i-1'. In the HNCO experiment, depicted in Figure 4.4 (p.96), magnetization is initially generated on $^1\text{H}_\text{N}$ [Step 1]. An INEPT pulse sequence transfers magnetization to ^{15}N via $^1J_{\text{NH}}$ (~90 Hz) coupling [Step 2] and all ^{15}N chemical shifts are recorded over the acquisition time t_1 [Step 3]. This experiment is optimized through the use of a constant time interval that allows overlap of the period needed to encode ^{15}N frequencies (t_1) and the period needed for dephasing of the ^{15}N spins with respect to the ^{13}CO . The introduction of this constant time period is advantageous as it results in a reduction of relaxation losses and a concomitant improvement in sensitivity. A second INEPT pulse sequence is used to transfer magnetization to the preceding residue ^{13}CO via $^1J_{\text{NCO}}$ (~15 Hz) coupling [Step 4] and all ^{13}CO chemical shifts are recorded over the acquisition period t_2 [Step 5]. Magnetization is then transferred back to $^1\text{H}_\text{N}$ using two reverse INEPT sequences [Step 6] and all $^1\text{H}_\text{N}$ chemical shift are collected over the third and final acquisition period t_3 [Step 7]. Fourier transformation of the resultant FID yields a frequency domain map with correlations between $^{13}\text{CO}(i-1)$, $^{15}\text{N}(i)$, and $^1\text{H}_\text{N}(i)$.

"Deuterium-modified" NMR experiments. Triple-resonance (^{15}N , ^{13}C , ^1H) experiments applied to proteins with molecular weights in excess of 20 kDa may experience a number of significant limitations (Gardner and Kay, 1998). The first problem relates to the rapid T_2 relaxation rates of nuclei (particularly ^{13}C) that participate in the magnetization transfer steps during the course of these

experiments. These rapid relaxation rates can lead to significant signal loss in these multiple step triple resonance experiments. The second problem, according to Gardner and Kay, (1998), is the lack of resolution in these spectra, particularly in the carbon dimension, due to the short acquisition times that must be used as a consequence of efficient ^{13}C T_2 relaxation rates. The ^{13}C T_2 relaxation rate is dominated by the strong dipolar interaction with its attached protons (Browne et al., 1973). As discussed in Chapter 1, replacing protons with deuterium, whose magnetogyric ratio is ~ 6.5 times lower than that of protons, greatly reduces the dipolar interaction and consequently lessens the ^{13}C T_2 relaxation rates (or increases ^{13}C T_2 relaxation times). This increase in ^{13}C T_2 relaxation times provides a solution to the above problems by reducing signal loss and increasing the time allotted for acquiring ^{13}C chemical shifts.

In general, the following experiments (CT-HNCA, CT-HN(CO)CA, CT-HN(CA)CB) have been modified by incorporating deuterium decoupling during periods in which transverse carbon magnetization is present (Gardner and Kay, 1998). At high magnetic field strengths, the ^2H spin lattice relaxation time, T_1 , in proteins is in the millisecond range due to the large ^2H quadrupolar interaction (~ 170 kHz) (Grzesiek et al., 1993). As a consequence, the ^2H - ^{13}C J coupling (~ 22 Hz) gives rise to a ^{13}C resonance that is broadened by scalar relaxation of the second kind (Abragam, 1961; London, 1990). However, incorporating deuterium decoupling into these pulse sequences during ^{13}C chemical shift acquisition effectively removes this broadening. This results in ^{13}C linewidths that are much narrower than those noted for protonated ^{13}C nuclei (Grzesiek et al., 1993).

3D "deuterium-modified" CT-HNCA. The deuterium-modified CT-HNCA (Yamazaki et al., 1994) was used to provide protein backbone information by correlating the $^1\text{H}_\text{N}$ with directly attached ^{15}N , $^{13}\text{C}_\alpha$ and preceding $^{13}\text{C}_{\alpha(i-1)}$ chemical shifts. In the HNCA experiment, shown in Figure 4.5 (p.97), magnetization is initially generated on the $^1\text{H}_\text{N}$ [Step 1]. An INEPT pulse sequence transfers magnetization to ^{15}N via the $^1\text{J}_{\text{NH}}$ (~ 90 Hz) coupling [Step 2]. A second

INEPT pulse sequence is used to transfer magnetization to the attached and preceding $^{13}\text{C}_\alpha$ via the $^1J_{\text{NC}\alpha}$ (8 – 12 Hz) and $^2J_{\text{NC}\alpha}$ (~ 7 Hz) couplings respectively [Step 3]. The two $^{13}\text{C}_\alpha$ chemical shifts are then recorded over the acquisition period t_1 [Step 4]. This experiment is optimized through the use of a constant time interval that allows overlap of the period needed to encode $^{13}\text{C}_\alpha$ frequencies (t_1) and the period needed for dephasing of the $^{13}\text{C}_\alpha$ spins with respect to the ^{15}N . This results in a reduction of relaxation losses and a concomitant improvement in sensitivity. Magnetization is then transferred back to ^{15}N using a reverse INEPT sequence [Step 5] and all ^{15}N chemical shifts are recorded over the acquisition time t_2 [Step 6]. Magnetization is then transferred back to $^1\text{H}_\text{N}$ using a reverse INEPT sequence [Step 7] and all $^1\text{H}_\text{N}$ chemical shifts are collected over the third and final acquisition period t_3 [Step 8]. Fourier transformation of the resultant FID yields a frequency domain map with correlations between $^{13}\text{C}_\alpha(i, i-1)$, $^{15}\text{N}(i)$, and $^1\text{H}_\text{N}(i)$.

3D “deuterium-modified” CT-HN(CO)CA. The deuterium modified 3D HN(CO)CA (Yamazaki et al., 1994) was used to complement the deuterium-modified HNCA by correlating $^1\text{H}_\text{N}$ chemical shifts with attached ^{15}N chemical shifts and preceding $^{13}\text{C}_\alpha(i-1)$ chemical shifts. In the HN(CO)CA experiment, shown in Figure 4.6 (p.98), magnetization is initially generated on $^1\text{H}_\text{N}$ [Step 1]. An INEPT pulse sequence transfers magnetization to ^{15}N via the $^1J_{\text{NH}}$ (~90 Hz) coupling [Step 2]. A second INEPT pulse sequence is used to transfer magnetization to the preceding residue $^{13}\text{CO}(i-1)$ via the $^1J_{\text{NCO}}$ (15 Hz) coupling [Step 3]. A third INEPT pulse sequence is used to transfer magnetization to the preceding residue $^{13}\text{C}_\alpha(i-1)$ via the $^1J_{\text{CaCO}}$ coupling (55 Hz) [Step 4]. The preceding $^{13}\text{C}_\alpha(i-1)$ chemical shifts are then recorded over the acquisition period t_1 [Step 5]. This experiment is optimized through the use of a constant time interval that allows overlap of the period needed to encode $^{13}\text{C}_\alpha$ frequencies (t_1) and the period needed for dephasing of the $^{13}\text{C}_\alpha$ spins with respect to the ^{13}CO . This results in a reduction of relaxation losses and a concomitant improvement in sensitivity. Magnetization is then transferred back to $^{15}\text{N}(i)$ using two reverse

INEPT sequence [Step 6] and all ^{15}N chemical shifts are recorded over the acquisition time t_2 [Step 7]. Finally, magnetization is transferred back to $^1\text{H}_\text{N}$ using a reverse INEPT sequence [Step 8] and all $^1\text{H}_\text{N}$ chemical shifts are collected over the third and final acquisition period t_3 [Step 9]. Fourier transformation of the resultant FID yields a frequency domain map with correlations between $^{13}\text{C}_\alpha(i-1)$, $^{15}\text{N}(i)$, and $^1\text{H}_\text{N}(i)$.

3D "deuterium modified" HN(CA)CB. The deuterium modified 3D HN(CA)CB (Yamazaki et al., 1994) is illustrated in Figure 4.7 (p.99). This experiment provides protein backbone information by correlating the $^1\text{H}_\text{N}$ chemical shifts with the attached ^{15}N chemical shifts and $^{13}\text{C}_\beta(i, i-1)$ chemical shifts. In this experiment, magnetization is first generated on $^1\text{H}_\text{N}$ [Step 1]. INEPT pulse sequences are used to transfer magnetization to ^{15}N via the $^1\text{J}_{\text{NH}}$ (~ 90 Hz) coupling [Step 2], the attached and preceding $^{13}\text{C}_\alpha(i, i-1)$ via the $^1\text{J}_{\text{NC}\alpha}$ (8 – 12 Hz) and $^2\text{J}_{\text{NC}\alpha}$ (~ 7 Hz) couplings [Step 3] and finally to the $^{13}\text{C}_\beta(i, i-1)$ [Step 4] via the $^1\text{J}_{\text{CC}}$ ($\sim 30\text{--}40$ Hz) coupling. The two $^{13}\text{C}_\beta(i, i-1)$ chemical shifts are then recorded over the acquisition period t_1 [Step 5]. Magnetization is then transferred back to ^{15}N using two reverse INEPT pulse sequences [Steps 6-7] and all ^{15}N chemical shifts are recorded over the acquisition time t_2 [Step 8]. Magnetization is finally transferred back to $^1\text{H}_\text{N}$ using a reverse INEPT sequence [Step 9] and all $^1\text{H}_\text{N}$ chemical shifts are recorded over the third and final acquisition period t_3 [Step 10]. Fourier transformation of the resultant FID yields a frequency domain map with correlations between $^{13}\text{C}_\beta(i, i-1)$, $^{15}\text{N}(i)$, and $^1\text{H}_\text{N}(i)$.

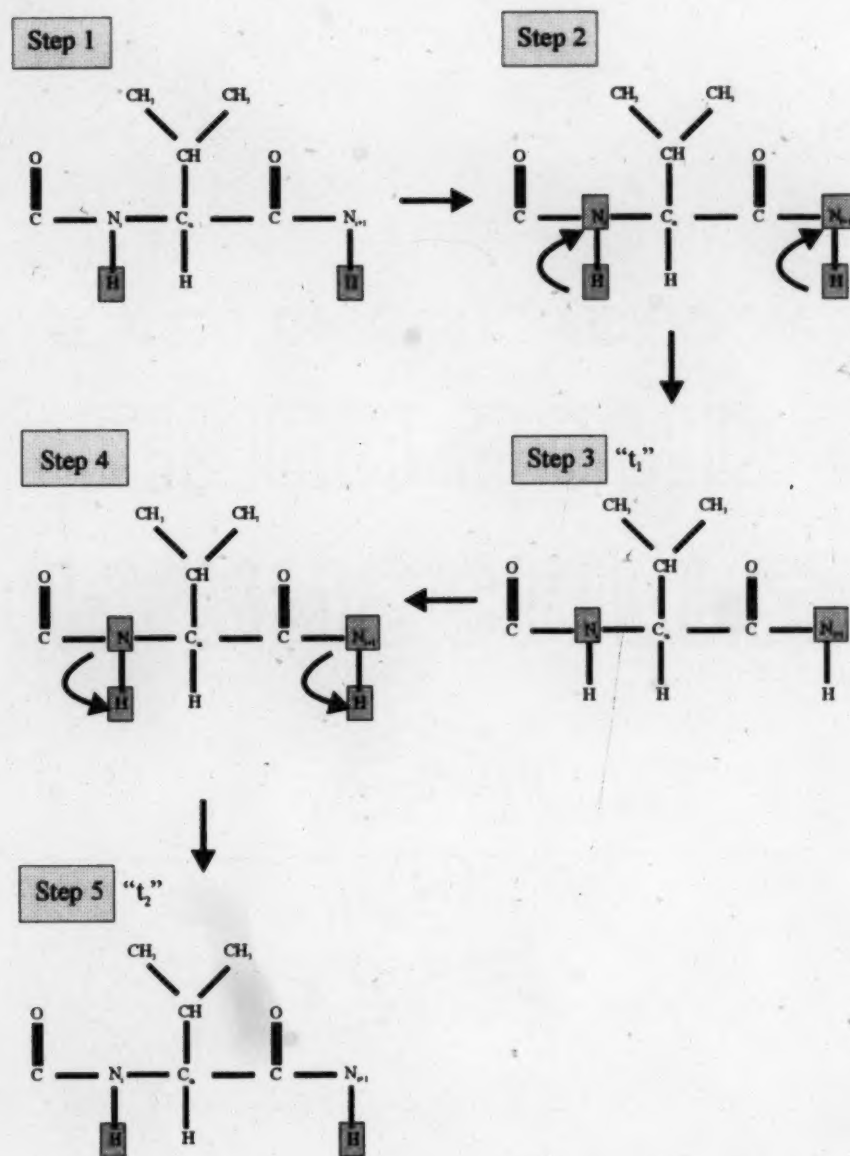


Figure 4.1. Outline of magnetization transfer in the ^1H - ^{15}N HSQC experiment.

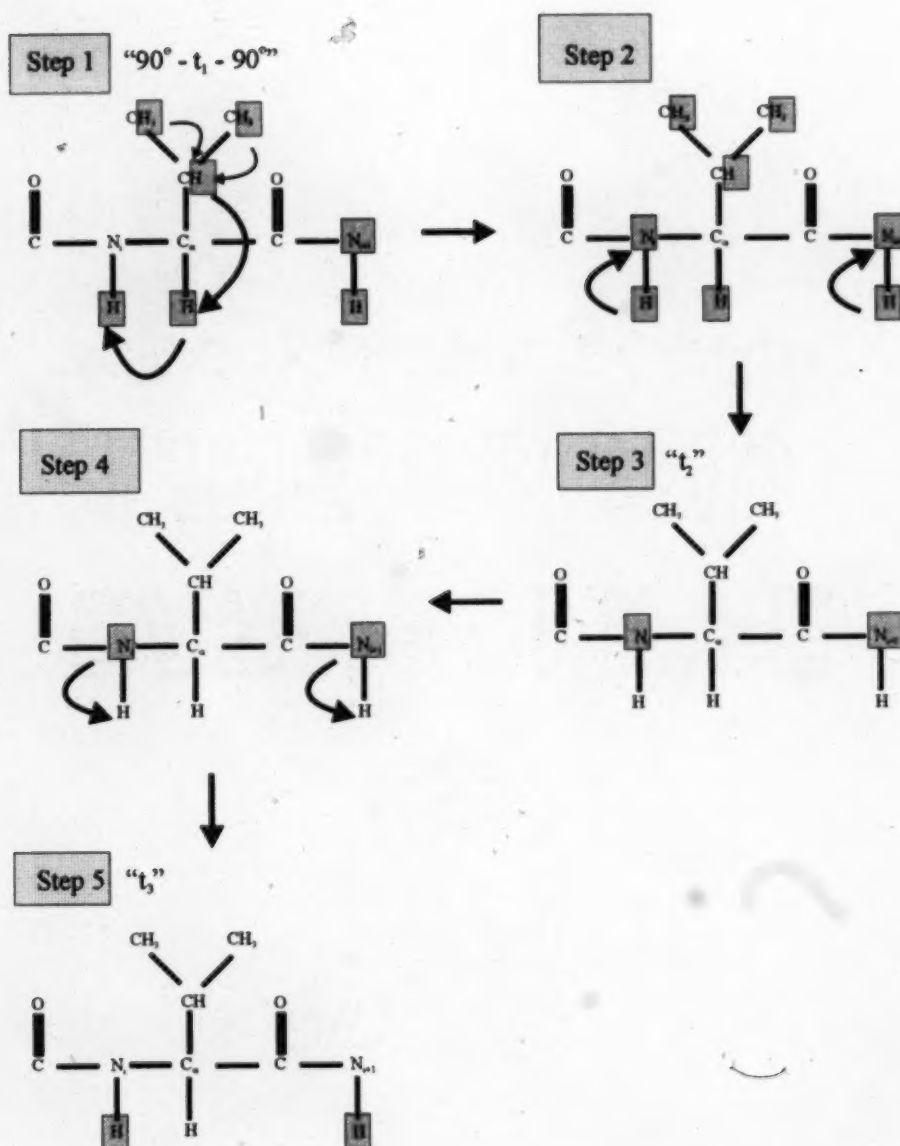


Figure 4.2. Outline of magnetization transfer in the ^1H - ^{15}N TOCSY-HSQC experiment.

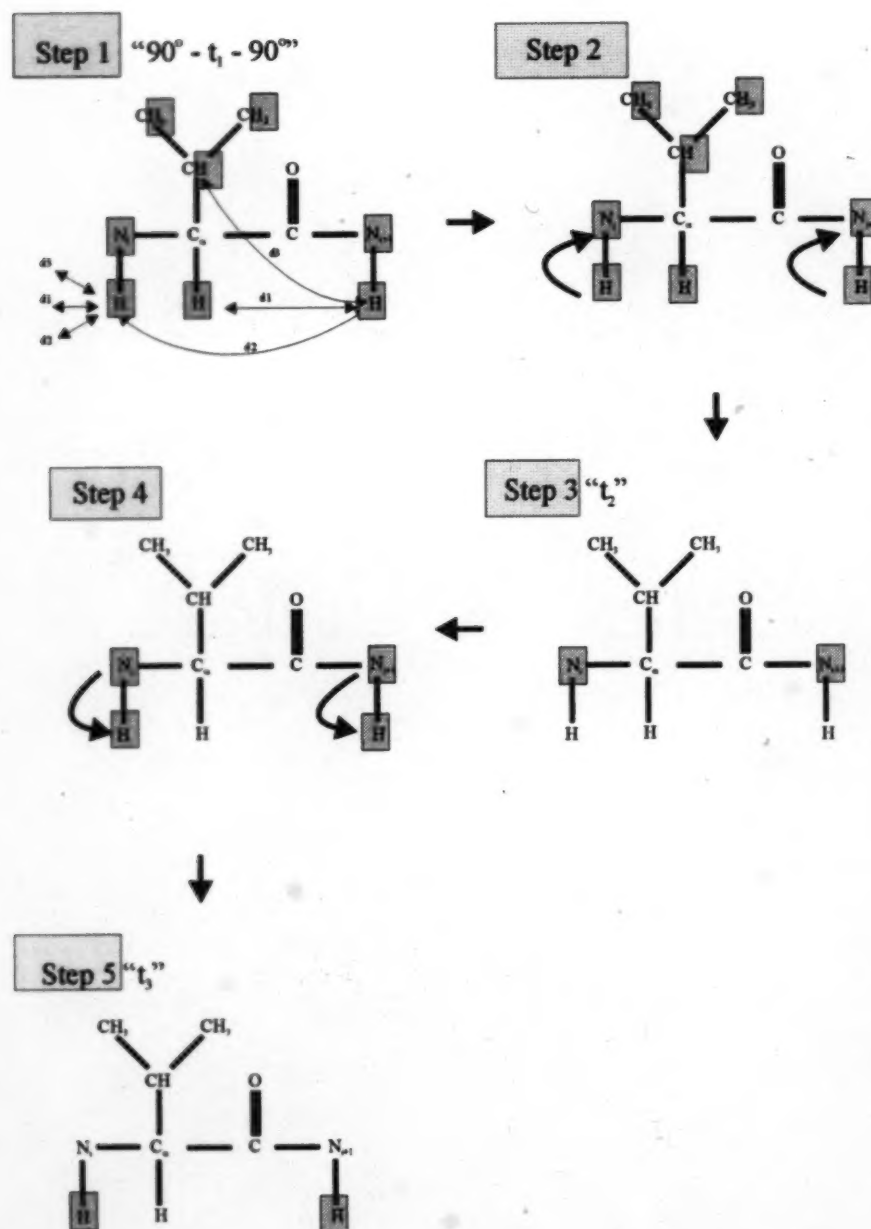


Figure 4.3. Outline of magnetization transfer in the ^1H - ^{15}N NOESY-HSQC experiment.

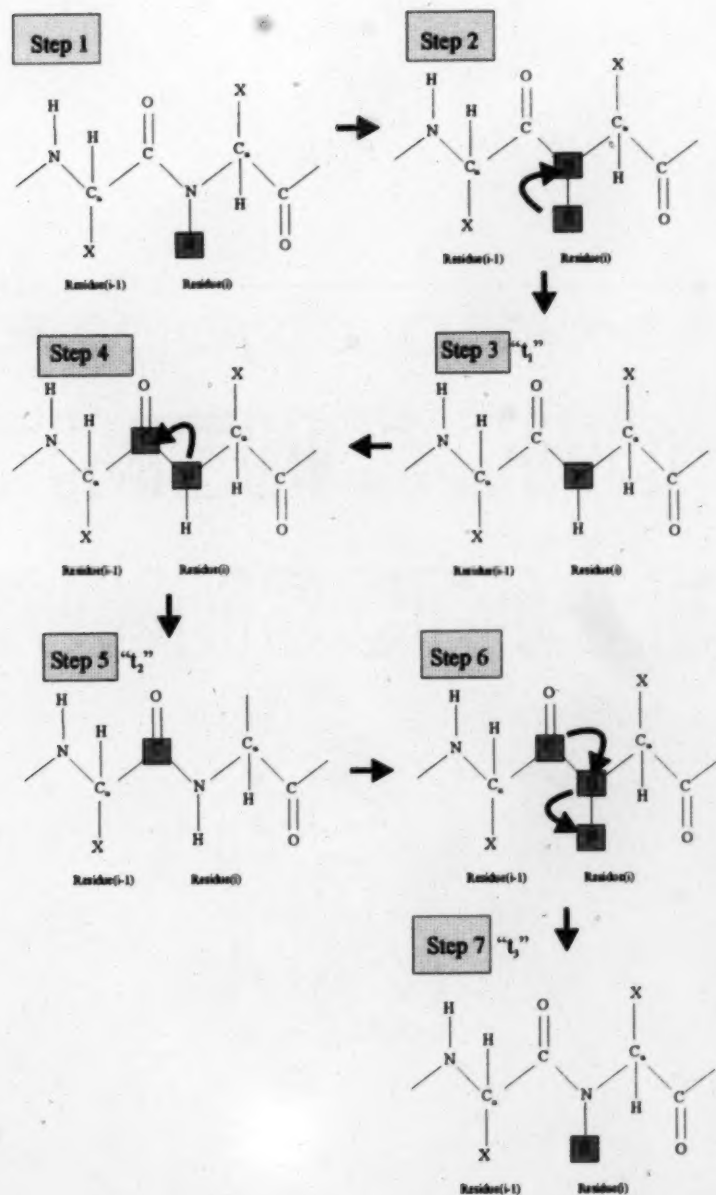


Figure 4.4. Outline of magnetization transfer in the CT-HNCO experiment.

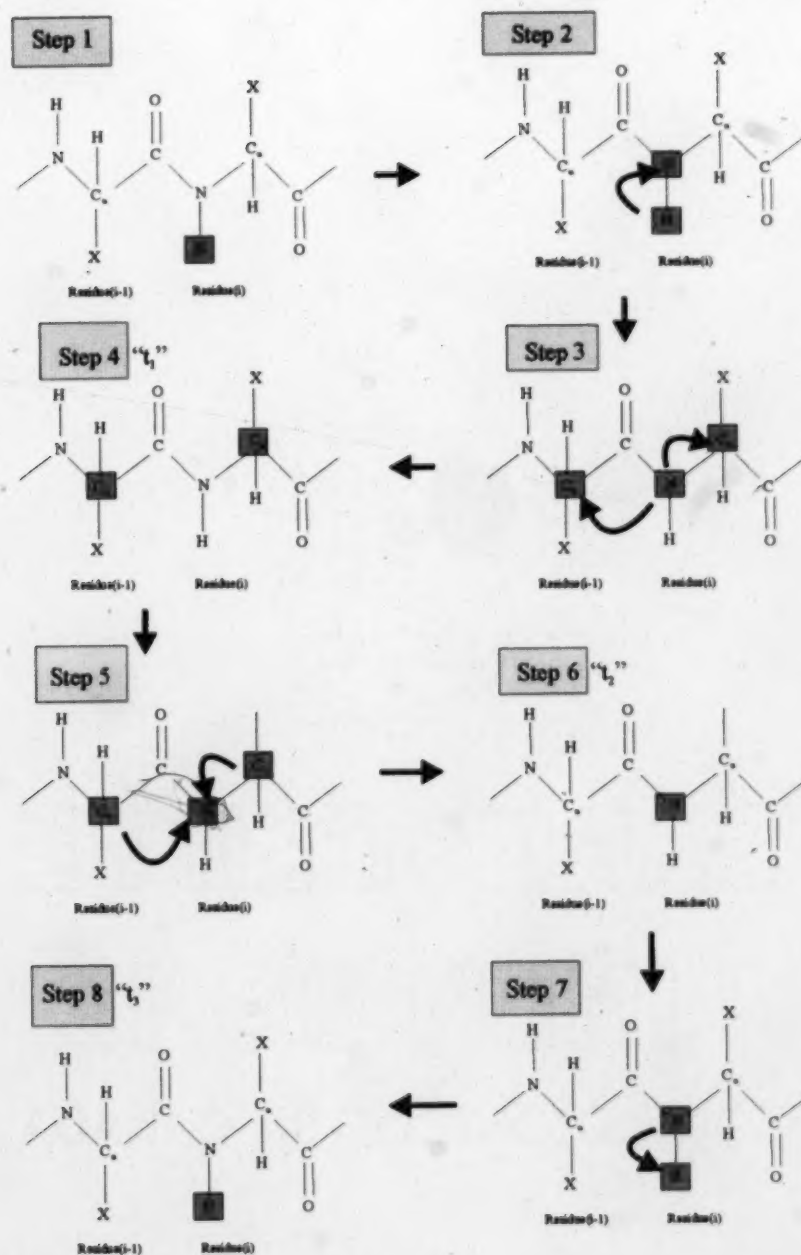


Figure 4.5. Outline of magnetization transfer in the "deuterium modified" HNCA experiment.

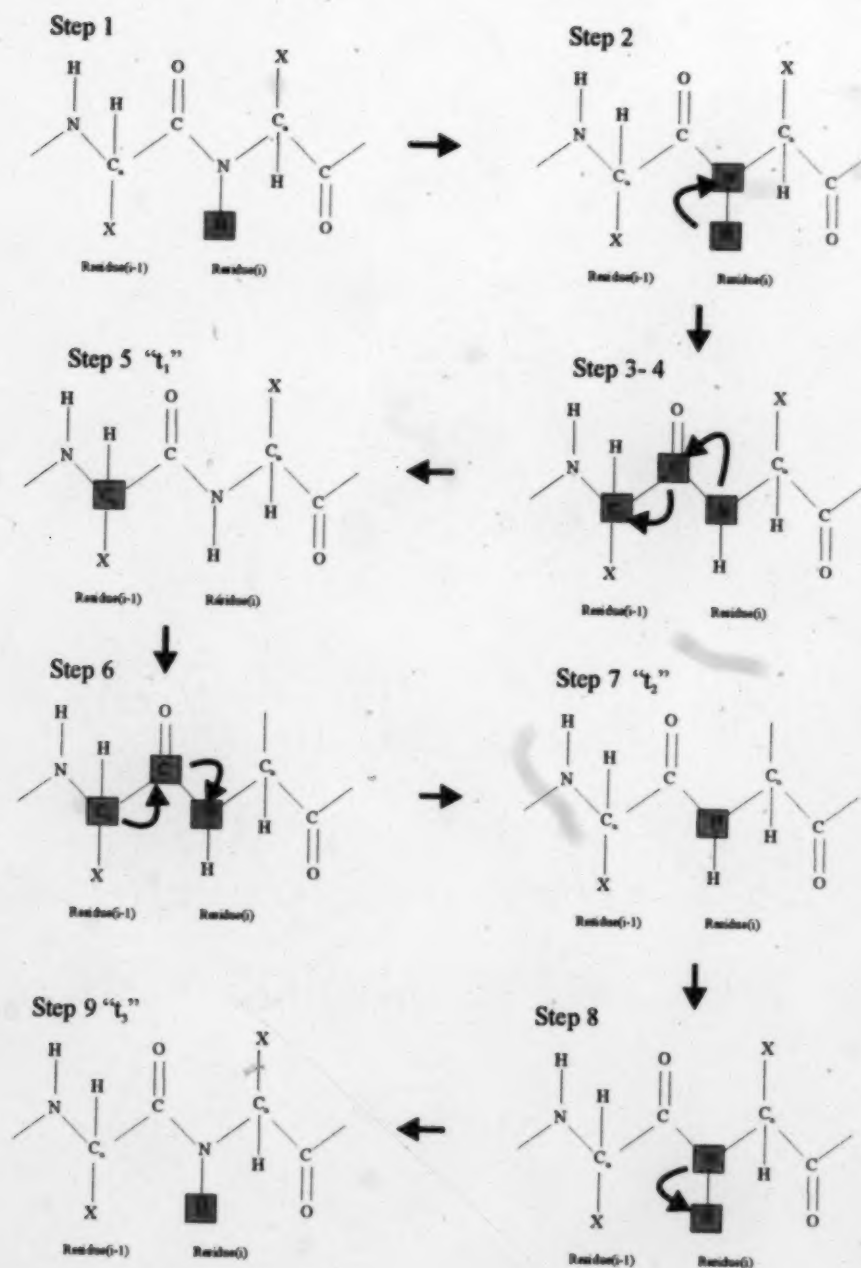


Figure 4.6. Outline of magnetization transfer in the "deuterium modified" HN(CO)CA experiment.

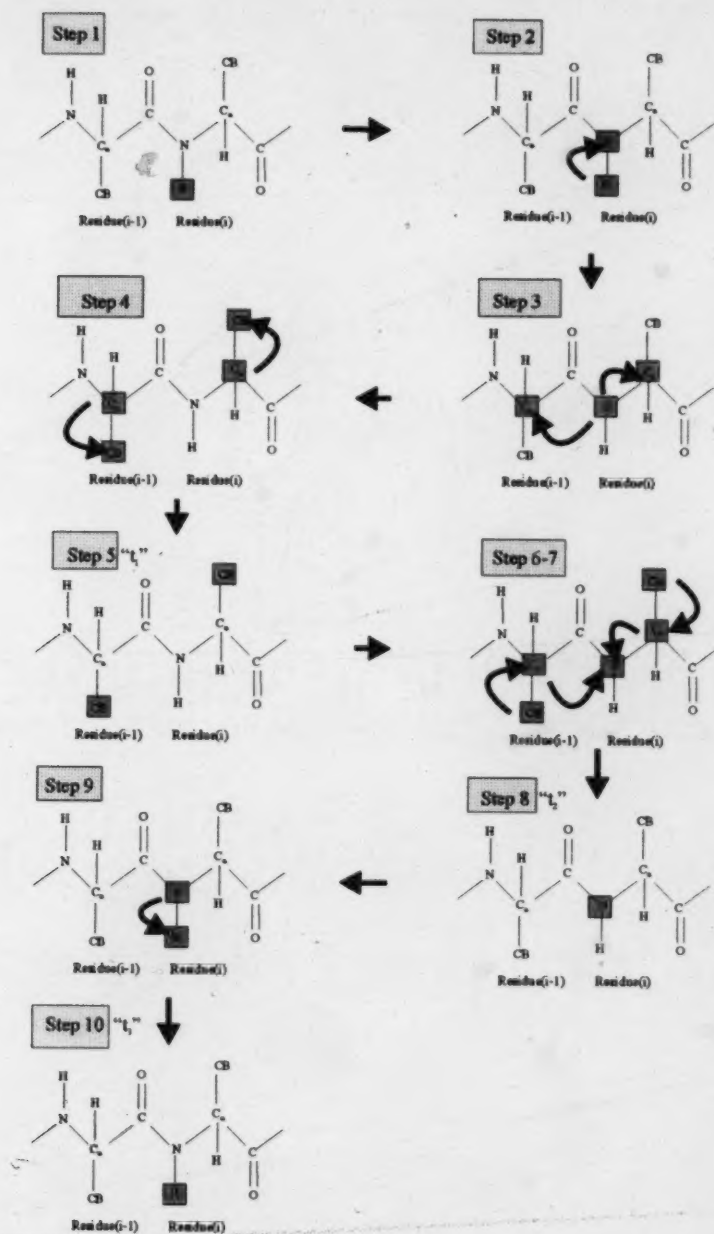


Figure 4.7. Outline of magnetization transfer in the "deuterium modified" HN(CA)CB experiment.

4.3 Materials and Methods

Preparation of HAV 3C protease NMR samples. ^{15}N , $^{13}\text{C}/^{15}\text{N}$, $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ and residue-specific ^{15}N labelled HAV 3C protease samples were isolated from the prototrophic *E. coli* strain MM294 containing the overexpression plasmid pHAV-3CEX (Malcolm et al., 1993) with a C24S mutation (to facilitate protein stabilization). Media was prepared, bacteria cultures were grown and the HAV 3C protease was isolated as described in Chapters 2 and 3. The NMR sample used for ^1H - ^{15}N HSQC, ^1H - ^{15}N TOCSY-HSQC and ^1H - ^{15}N NOESY-HSQC experiments contained 0.75 mM HAV 3C protease (uniformly ^{15}N labelled) dissolved in 350 mM NaCl, 20 mM KH_2PO_4 , 0.5 mM EDTA at pH 7.0. The sample was prepared to a volume of 500 μL by combining 10 μL DSS, 100 μL D_2O (99%) and 390 μL concentrated protease solution. The resulting sample had a $\text{H}_2\text{O}:\text{D}_2\text{O}$ ratio of 8:2. The NMR sample used for the CT-HNCO experiment contained 1.0 mM HAV 3C protease (uniformly $^{13}\text{C}/^{15}\text{N}$ labelled) dissolved in 350 mM NaCl, 20 mM KH_2PO_4 , 0.5 mM EDTA at pH 7.0. The sample was prepared to a volume of 500 μL by combining 10 μL of 3% (w/v) sodium azide, 10 μL DSS, 50 μL D_2O (99%) and 430 μL concentrated protease solution. The resulting sample had a $\text{H}_2\text{O}:\text{D}_2\text{O}$ ratio of 9:1. The NMR sample used for the deuterium-modified CT-HNCA, CT-HN(CO)CA and HN(CA)CB experiments contained 1.0 mM HAV 3C protease (uniformly $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ labelled) dissolved in 100 mM NaCl, 20 mM KH_2PO_4 , 0.5 mM EDTA, 1.0 mM DTT at pH 5.4. The sample was prepared to a volume of 500 μL by combining 10 μL of 3% (w/v) sodium azide, 10 μL DSS, 50 μL D_2O (99%) and 430 μL concentrated protease solution. The resulting sample had a $\text{H}_2\text{O}:\text{D}_2\text{O}$ ratio of 9:1.

Experimental set up for NMR experiments. Experiments were set up and performed by Dr. Stephane Gange, Dr. Carolyn Slupsky, Dr. Leo Spyropoulos and Dr. Brian Sykes. ^1H - ^{15}N HSQC, ^1H - ^{15}N TOCSY-HSQC and ^1H - ^{15}N NOESY-HSQC experiments were carried out at 25°C on a Varian Unity 600 MHz spectrometer equipped with three channels and a pulse-field gradient triple

Table 4.1. Acquisition parameters for NMR experiments.

Experiment	Nucleus			No. of complex points			Spectral widths (Hz)			Scans
	F1	F2	F3	F1	F2	F3	F1	F2	F3	
^1H - ^{15}N HSQC (pH 7.0)	^{15}N	$^1\text{H}^{\text{N}}$		256	1024		2000	8000		224
^1H - ^{15}N TOCSY HSQC	^1H	^{15}N	$^1\text{H}^{\text{N}}$	240	64	1024	7000	1885	8000	12
^1H - ^{15}N NOESY HSQC	^1H	^{15}N	$^1\text{H}^{\text{N}}$	256	64	1024	7000	1885	8000	16
HNCO	^{15}N	^{13}CO	$^1\text{H}^{\text{N}}$	64	128	1024	1625	1541	6000	24
^1H - ^{15}N HSQC (pH 5.4)	^{15}N	$^1\text{H}^{\text{N}}$		256	1024		1600	8000		32
HNCA	$^{13}\text{C}_{\alpha}$	^{15}N	$^1\text{H}^{\text{N}}$	168	56	1024	3268	1600	8000	16
HN(CO)CA	$^{13}\text{C}_{\alpha}$	^{15}N	$^1\text{H}^{\text{N}}$	168	56	1024	3268	1600	8000	16
HN(CA)CB	$^{13}\text{C}_{\beta}$	^{15}N	$^1\text{H}^{\text{N}}$	128	56	1024	7650	1600	8000	16

resonance probe. Mixing times in the ^{15}N - ^1H TOCSY-HSQC and ^{15}N - ^1H NOESY-HSQC experiments were 32 ms and 72 ms respectively. All triple resonance experiments were acquired at 25°C on a Varian Unity 500 MHz spectrometer equipped with three channels and a pulsed-field gradient triple resonance probe (with sequential quadrature detection during the detection period). The acquisition parameters for all experiments are listed in Table 4.1 (p.101). Chemical shifts were directly (^1H) and indirectly (^{13}C , ^{15}N) referenced to DSS at 0 ppm as described by Wishart et al. (1995). Spectra were processed and peak-picked on a Sun Ultra 1 workstation, using the NMRPipe (Delaglio et al., 1995) and PIPP (Garrett et al., 1991) software packages.

Assignment strategy. A two-step approach was adopted to achieve the sequential polypeptide backbone resonance assignments. The first step was to use the ^1H - ^{15}N HSQC (Figures 4.8 and 4.9, p.104-105) spectrum of HAV 3C protease to make unambiguous assignments for a number of amino acid types and use them as starting and check-points in later data tracing. This required the use of ^1H - ^{15}N HSQC spectra from residue-specific ^{15}N labelled HAV 3C protease samples (as discussed in Chapter 3). Using these spectra, $^1\text{H}_\text{N}$ and ^{15}N assignments were made for a significant portion of the alanine, glycine, leucine, lysine, isoleucine and valine residue types. It is important to note that all ^1H - ^{15}N HSQC spectra for the residue-specific ^{15}N labelled HAV 3C protease samples, except the isoleucine sample, were collected at pH 7.0 whereas the ^1H - ^{15}N HSQC spectra of uniform and isoleucine ^{15}N -labelled HAV 3C protease were collected at pH 5.4. This was somewhat problematic due to differences in the amide chemical shifts observed at different pH values. Only those resonances in the residue-specific HSQC spectra that could be unambiguously assigned to peaks in the uniformly labelled HSQC spectrum were used.

The second step in the assignment process was to find sequential connectivities of the backbone resonances. Initially, chemical shift "clusters" were assembled by comparing spin system data collected from HNCA, HN(CO)CA,

HN(CA)CB and the residue-specific HSQC data. This process began by using the residue-specific HSQC data to assign a residue type to a peak in the uniformly labelled HSQC spectrum. Next its $^1\text{H}_\text{N}$, ^{15}N , $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts ('i' chemical shifts) would be determined along with the $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts of the preceding residue ('i-1' chemical shifts). This 'spin system' assignment process was repeated for all peaks in the uniformly labelled HSQC spectrum. Chemical shift clusters were built in the 'i-1' direction by correlating the $^{13}\text{C}_\alpha(\text{i})$ and $^{13}\text{C}_\beta(\text{i})$ chemical shifts from one spin system with the original (known residue-type) spin system's $^{13}\text{C}_\alpha(\text{i-1})$ and $^{13}\text{C}_\beta(\text{i-1})$ chemical shifts. Similarly, clusters could be built moving in the 'i + 1' direction by reversing the above protocol. This process is demonstrated in Figure 4.10 (p.106), which shows sequential connectivities being made between G195 to L199 using information afforded by the HNCA spectrum. When a sizeable cluster was built (4-6 residues long), we would observe what residue type we started with (eg. Lys) and observe the other spin systems involved in the cluster. Usually, one of the other spin systems would also have been assigned to a specific residue type. Additionally, ideas about what residues were responsible for the other spin systems in the cluster could be generated by comparing the observed chemical shifts of the spin system to expected random coil chemical shifts for residue-type spin systems (Wishart et al., 1995). Once this was accomplished for a given cluster, the string of residues (known residue types and hypothesized residue types) was then fitted into the HAV 3C sequence. When a positive match for a given sequential cluster could be made, the sequential backbone chemical shift information was inserted into Table 4.2 (p.109).

Because of the large size of the HAV 3C protease (23.8 kDa), chemical shift degeneracy often posed a challenging problem. In the process of either building up one cluster or linking two clusters together, it was always necessary to match up one or two chemical shifts (^{13}C and ^{13}C). In some cases, multiple choices were available. When a positive decision could not be made, we would stop and start from another marker, that being of a known residue type.

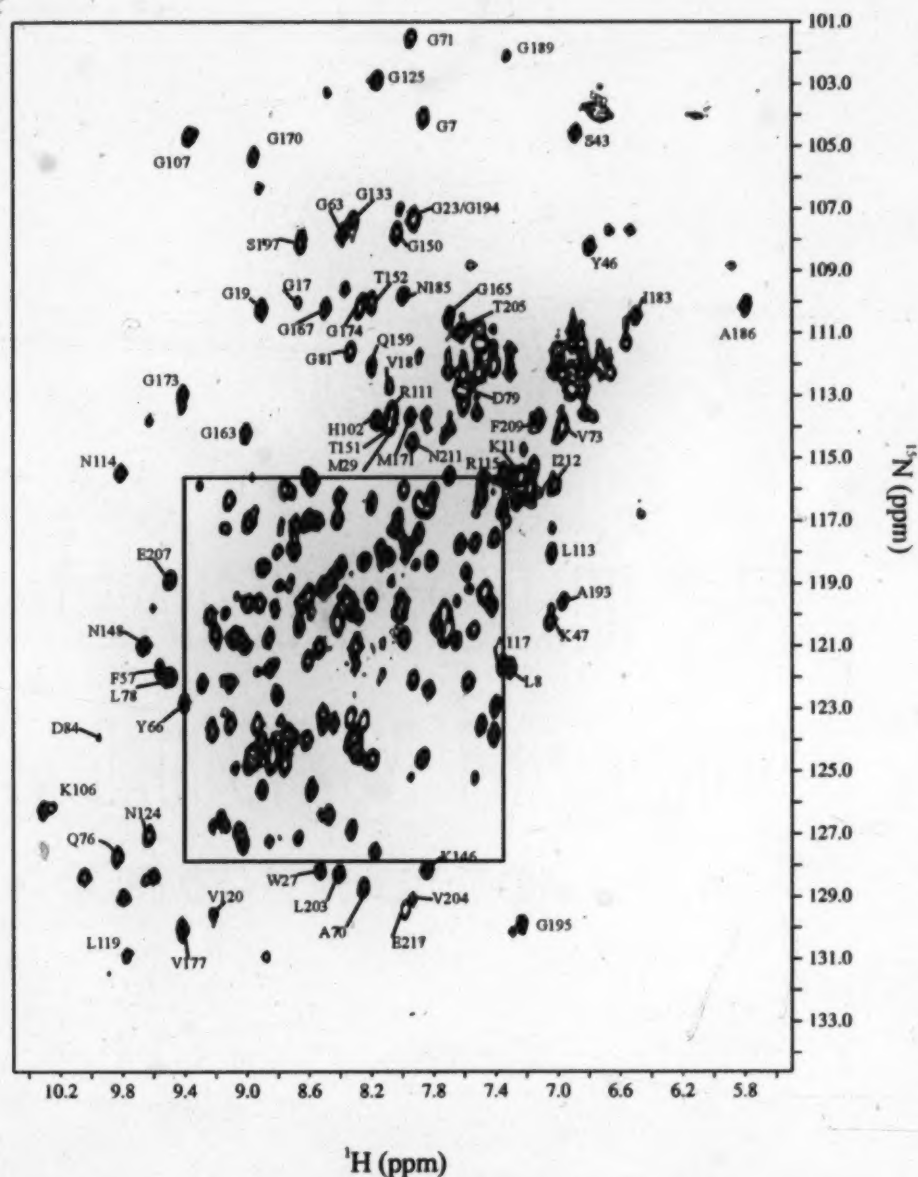


Figure 4.8. ^1H - ^{15}N HSQC spectrum from 1 mM HAV 3C protease in 20 mM KH_2PO_4 , 100 mM NaCl, 0.5 mM EDTA, 1.0 mM DTT at pH 5.4 collected at 25°C on a Varian Unity 600 MHz spectrometer.

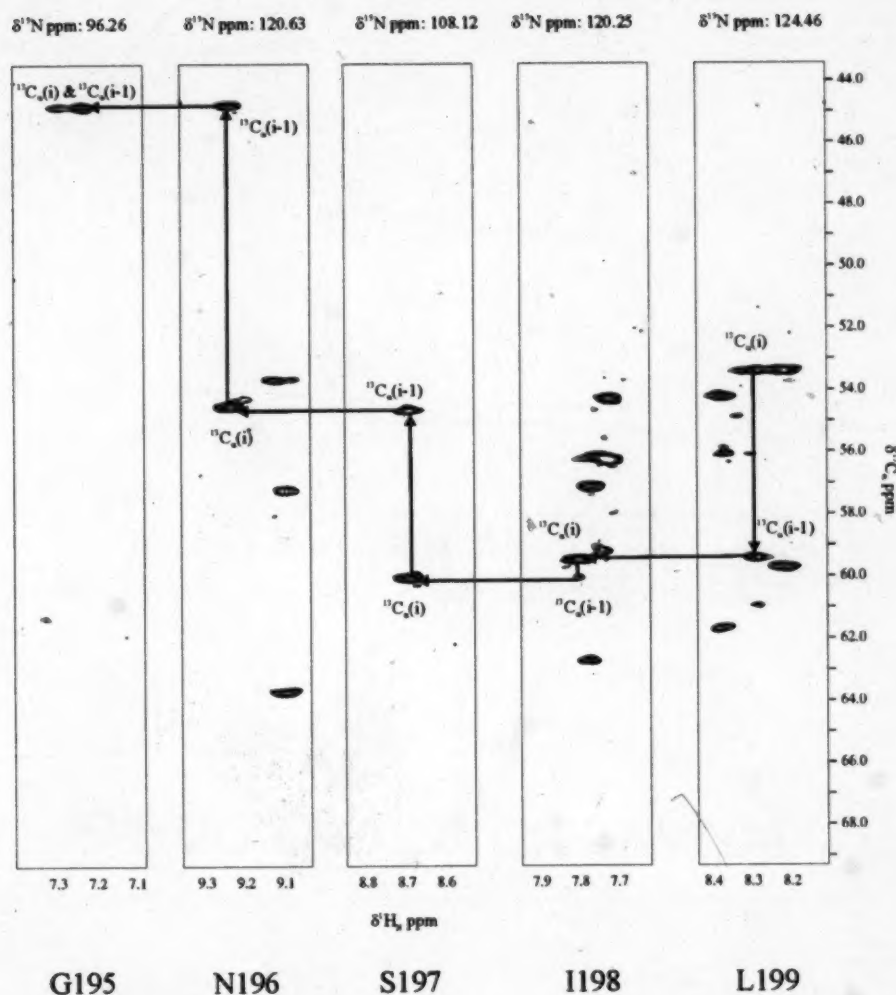


Figure 4.10. Polypeptide backbone assignments of Gly195, Asn196, Ser197, Ile198 and Leu199. An example how assignments could be extended in both directions using the 3D HNCA. The HNCA experiment was performed on the ^2H , ^{13}C , ^{15}N uniformly labelled HAV 3C protease (1mM) in 20 mM KH_2PO_4 , 100 mM NaCl, 0.5 mM EDTA, 1.0 mM DTT, 10% D_2O , pH 5.4 at 25°C using a Varian Unity 500 MHz spectrometer.

4.4 Results and Discussion

Spin system assignments. The general approach for assigning the backbone spin system resonances ($^1\text{H}_\text{N}$, ^{15}N , $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, $^1\text{H}_\alpha$ and $^{13}\text{CO}(\text{i}-1)$) was based on the use of the ^1H - ^{15}N HSQC spectrum collected at pH 5.4 (Figure 4.8) as a chemical shift fingerprint of the observable amides. This spectrum displayed 204 of the expected 210 resonances (217 – 7 prolines). This less-than-perfect result is due to several amide resonances having the same (or very similar) $^1\text{H}_\text{N}$ and ^{15}N chemical shifts. Amide HN pairs identified in the HSQC were correlated to the backbone ^{13}C resonances using HNCA and HN(CO)CA spectra. The HNCA spectrum provided both intraresidue (i) and interresidue (i-1) $^{13}\text{C}_\alpha$ resonances with the intraresidue resonance usually being of a noticeably higher intensity. The HN(CO)CA spectrum was used to unambiguously determine which resonance in the HNCA spectrum belonged to the interresidue $^{13}\text{C}_\alpha$ resonance. Using these two spectra, 205 $^{13}\text{C}_\alpha$ resonances could be identified. Similarly, HN pairs were correlated to $^{13}\text{C}_\beta$ resonances using the HN(CA)CB spectrum. This spectrum provided both intraresidue and interresidue $^{13}\text{C}_\beta$ resonance information. The intraresidue $^{13}\text{C}_\beta$ resonance was assumed to be the more intense of the two. The HN(CA)CB experiment was not as sensitive as the HNCA and HN(CO)CA experiments and consequently only 170 $^{13}\text{C}_\beta$ resonances could be identified.

Determining $^1\text{H}_\alpha$ and $^{13}\text{CO}(\text{i}-1)$ resonances for each spin system required the use of ^1H - ^{15}N HSQC, ^1H - ^{15}N TOCSY-HSQC, ^1H - ^{15}N NOESY-HSQC, and HNCO spectra, all of which were acquired at pH 7.0. This was done after the sequential assignments had been made using the HNCA, HN(CO)CA, HN(CA)CB and residue-specific HSQC spectra. To determine $^1\text{H}_\alpha$ and $^{13}\text{CO}(\text{i}-1)$ resonance assignments, it was necessary to determine which HN resonances in the 'high-pH' ^1H - ^{15}N HSQC could be unambiguously equated to HN resonances in the 'low-pH' ^1H - ^{15}N HSQC. Differences in the amide chemical shifts between these two spectra likely arose from differences in protonation states of acidic (aspartate, glutamate, cysteine and tyrosine) and basic (arginine, lysine, histidine) amino acids.

Changing the protonation-state will invariably change the surrounding electronic environment which amide chemical shifts are notoriously sensitive to (MacLachlan et al., 1997). Consequently, with so many differing amide resonances, equivalency could only be established for 71 HN pairs between the 'high-pH' and 'low-pH' HSQC spectra.

The equivalent, 'high-pH' HN pairs were subsequently correlated to ^1H resonances using ^1H - ^{15}N TOCSY-HSQC and ^1H - ^{15}N NOESY-HSQC spectra. It is important to note that these spectra were of much lower quality than the other triple resonance spectra. If a ^1H chemical shift was not present in the TOCSY spectrum, the NOESY spectrum was used to determine the most likely $^1\text{H}_\alpha$ chemical shift based on the NOEs observed for that spin system (i.e. presence detected in the 'i + 1' spin system). Similarly, the equivalent HN pairs (at high pH) were correlated to $^{13}\text{CO}(i-1)$ resonances using the HNCO spectrum. Because the HNCO was a triple resonance experiment, this spectrum was significantly more sensitive than the ^{15}N -edited experiments (TOCSY and NOESY) and nearly all expected ^{13}CO resonances could be observed. However, only those $^{13}\text{CO}(i-1)$ chemical shifts that correlated to the 71 HN pairs are reported. The last step was to assign the $^{13}\text{CO}(i-1)$ resonances to their own respective (i) spin system according to the HAV 3C protease sequence.

Sequential assignments. Chemical shift information from HNCA, HN(CO)CA, HN(CA)CB and information from the ^1H - ^{15}N HSQC spectra of the ^{15}N residue-specific labelled proteases have allowed 175 sequential assignments to be made (Table 4.2). Initially, HNCA, HN(CO)CA and residue-specific HSQC spectra were used to link up $^{13}\text{C}_\alpha(i)$ and $^{13}\text{C}_\alpha(i-1)$ resonances (Figure 4.10). However, many ambiguities arose due to similarities in the $^{13}\text{C}_\alpha$ resonances of amino acids such as histidine, leucine, methionine and glutamine. All of these residues typically have $^{13}\text{C}_\alpha$ random coil chemical shifts at approximately 53 ppm (Wishart et al., 1995b). Using this method, about 65 percent of the sequential assignments could be made. The $^{13}\text{C}_\beta$ resonance information collected from the HN(CA)CB

Table 4.2. Sequential backbone chemical shifts assignments of HAV 3C protease.

Residue	^{15}N	$^1\text{H}_\text{N}$	$^{13}\text{C}_\alpha$	$^{13}\text{C}_\beta$	$^{15}\text{N}^*$	$^1\text{H}_\text{N}^*$	$^1\text{H}_\alpha^*$	$^{13}\text{CO}^*$
1	S		56.89	63.69				
2	T	119.60	8.99	66.49	67.75			
3	L	119.46	8.19	57.33	40.38			
4	E	120.75	7.73	58.66	28.79			
5	I	119.71	7.41	61.01	34.25			178.54
6	A	122.15	9.10	55.08	17.60	122.26	9.10	3.92
7	G	104.03	7.86	46.68		104.34	7.85	3.92
8	L	121.68	7.30	57.12	40.78	122.31	7.34	4.05
9	V	118.24	8.24	66.16	30.93	118.68	8.24	3.35
10	R	116.91	8.60	59.51	28.61	117.1	8.58	3.83
11	K	115.62	7.29	57.75	31.41	116.04	7.32	4.25
12	N	118.63	7.58	51.96	38.32	118.99	7.60	5.03
13	L	125.20	7.53	55.40	42.66	125.45	7.55	4.98
14	V	117.25	9.17	58.29		117.50	9.11	5.17
15	Q	118.25	7.81	56.39	30.93	118.30	7.76	5.14
16	F		54.0					
17	G	109.99	8.67	43.64				169.59
18	V	112.74	8.08	58.56	35.01	112.33	8.12	5.25
19	G	110.22	8.90	45.32				
20	E	120.24	8.46	54.59	30.93			
21	K	123.48	8.93	58.04	31.02			
22	N	116.99	8.55	53.55	37.12			175.18
23	G	107.32	7.94	44.30		107.60	7.92	4.23
24	S	116.01	7.99	57.46	63.94	116.20	8.04	4.40
25	V	122.05	7.93	61.13	32.02	122.32	7.93	3.46
26	R	127.54	8.18	53.73	30.14	128.11	8.25	4.36
27	W	128.14	8.52	57.02	28.05			
28	V	120.68	9.07	63.55	34.33			
29	M	113.6	8.06	56.04	29.43			
30	N	119.09	8.51	52.55	42.03			
31	A	117.25	9.14	50.73	22.64			
32	L							
33	G							
34	V							
35	K							
36	D							
37	D							
38	W							
39	L							
40	L							
41	V							
42	P							
43	S							
44	H	119.98	9.23	57.73	26.01			
45	A	122.11	7.57	54.22	19.29			
46	Y	108.20	6.79	56.63	41.35			
47	K	120.22	7.04	57.27	31.1			
48	F	115.75	8.57	56.97	37.28			
49	E	119.23	7.46	54.15	28.86			
50	K	124.78	8.75	57.03	31.69			
51	D	119.64	8.91	54.68	38.72			

Residue	^{15}N	$^1\text{H}_\text{N}$	$^{13}\text{C}_\alpha$	$^{13}\text{C}_\beta$	$^{15}\text{N}^*$	$^1\text{H}_\text{N}^*$	$^1\text{H}_\alpha^*$	$^{13}\text{CO}^*$
52	Y	117.69	7.53	60.03	37.27			
53	E	118.42	8.39	58.19	27.45			
54	M	115.54	7.69	54.30	31.83			
55	M							
56	E							
57	F	121.72	9.56	57.33	39.88			
58	Y	123.8	8.74	55.88	39.90			
59	F	115.88	9.30	55.09	42.46			
60	N	120.99	9.01	51.45	38.55			
61	R	127.30	9.02	53.99	31.19			
62	G	116.34	9.11	46.79				174.95
63	G	107.79	8.38	44.59	108.00	8.41	4.01	173.87
64	T	117.52	7.40	61.40	117.72	7.41	4.20	
65	Y	124.07	8.33	53.93	37.43			
66	Y	122.77	9.40	57.17	40.58			
67	S	116.00	8.75	55.89	66.79			
68	I	120.73	8.86	59.50	41.02			
69	S	121.04	8.30	58.30	62.94			
70	A	128.67	8.24	54.59	18.40			
71	G	101.48	7.94	45.35				
72	N	116.55	7.87	52.26	39.33			175.08
73	V	114.02	6.97	60.35	32.43	114.75	7.00	4.56
74	V	124.57	7.87	62.13	32.42	125.14	7.89	4.25
75	I	126.37	8.47	58.80	39.40	126.84	8.49	5.27
76	Q	127.72	9.83	53.42	31.59	128.09	9.84	4.94
77	S	119.41	8.60	57.77	64.24			
78	L	121.97	9.49	55.93	41.86			
79	D	112.82	7.60	51.51	42.43			
80	V	118.47	8.89	62.50	31.73			
81	G	111.59	8.33	43.85				
82	F	121.47	8.60	58.46	38.80			
83	Q	122.38	7.83	52.97	27.85			175.50
84	D	123.93	9.95	53.70	123.86	9.95	5.51	
85	V	124.6	8.83	63.02				178.20
86	V	118.89	8.45	58.66	119.41	8.50	4.86	
87	L							
88	M							
89	K							
90	V							
91	P							
92	T							
93	I							
94	P							
95	K							
96	F							
97	R			55.88	29.69			
98	D	123.16	8.51	53.24	39.58			
99	I	121.81	8.93	60.88				177.89
100	T	117.06	8.98	65.22	68.85	117.02	8.99	3.60
101	Q	116.21	8.40	56.24	26.70	116.24	8.42	4.20
102	H	113.75	8.16	55.39	27.60			
103	F							
104	I			59.88	39.61			
105	K	126.89	9.05	55.93	31.90			179.34

Residue	¹⁵ N	¹ H _N	¹³ C _α	¹³ C _β	¹⁵ N*	¹ H _N *	¹ H _α *	¹³ CO*
106 K	126.30	10.3	60.01	31.30	126.56	10.33		179.86
107 G	104.65	9.37	45.48		104.88	9.43	3.94	174.35
108 D	117.20	8.04	53.99	41.83	117.49	8.07	5.06	177.05
109 V	122.92	7.40	68.65	30.00	123.17	7.42	3.93	
110 P			65.10					
111 R	113.49	8.06	56.97					
112 A	119.59	7.99	51.09	20.88				177.85
113 L	118.07	7.03	55.29	42.24	118.27	7.04	4.10	176.71
114 N	115.46	9.81	54.53	36.95	115.75	9.85	4.26	173.30
115 R	115.47	7.34	52.54	31.64	115.77	7.37	4.83	176.00
116 L	120.24	8.42	55.09	40.37	120.64	8.45	4.73	
117 A	121.56	7.35	51.31	23.78				
118 T	116.00	8.75	61.12	71.31				172.36
119 L	130.88	9.77	53.78	40.46	131.19	9.77	5.66	173.76
120 V	129.51	9.21	61.06	30.79	129.90	9.25	4.35	
121 T	118.15	8.08	58.31	68.86				
122 T	115.56	8.61	60.61	71.11				
123 V	117.81	7.95	61.73	32.74				
124 N	127.05	9.63	53.67	36.98				175.18
125 G	102.87	8.15	44.64		103.09	8.18	4.25	174.41
126 T	120.70	7.98	60.23	69.33	120.97	8.00	4.60	
127 P								
128 M								
129 L								
130 I								
131 S	120.02	7.73	58.90					
132 E	123.44	8.44	55.63	28.45				
133 G	107.36	8.32	44.80					
134 P			62.76	31.48				
135 L	119.73	8.32	54.59	36.87				
136 K			53.75					
137 M	126.50	9.16	53.45					
138 E	126.76	9.13	53.45	29.94				
139 E	124.6	8.83	59.24	28.71				
140 K	117.40	7.89	54.63	35.10				
141 A	124.36	8.75	50.63	21.96				
142 T	116.92	8.41	60.02	70.47				
143 Y			54.82	38.55				
144 V	116.38	8.19	60.63	33.66				
145 H	124.28	8.83	54.76	32.24				
146 K	128.13	7.84	54.87	32.10				176.76
147 K	123.99	8.61	55.84	32.47	124.36	8.70	4.51	178.22
148 N	121.00	9.66	55.49	37.18	122.11	9.63	4.60	
149 D	116.39	7.89	52.82	39.35				
150 G	107.77	8.03	44.90					174.62
151 T	113.92	8.08	61.65	69.98	114.05	8.08	4.47	173.50
152 T	110.00	8.19	59.87	71.03	109.66	8.17	5.28	174.53
153 V	119.59	8.64	59.47	34.16	119.76	8.74	4.31	172.95
154 D	124.60	8.19	53.13	40.40	125.26	8.13	4.89	175.15
155 L	125.58	8.90	53.22	41.87	125.96	8.89	4.43	
156 T	116.06	7.80	60.62	71.04				
157 V	117.11	8.68	57.89	34.43				
158 D			53.45	42.23				
159 Q	112.02	8.19	57.40	26.37				172.26

Residue		^{15}N	$^1\text{H}_\text{N}$	$^{13}\text{C}_\alpha$	$^{13}\text{C}_\beta$	$^{15}\text{N}^*$	$^1\text{H}_\text{N}^*$	$^1\text{H}_\alpha^*$	$^{13}\text{CO}^*$
160	A	126.84	8.32	49.43	21.57	126.54	8.31	4.98	
161	W	123.75	9.22	55.17	31.14				
162	R	122.18	9.28	54.44	33.03				
163	G	114.20	9.00	43.15					172.35
164	K	122.55	8.80	55.14	33.82	122.83	8.83	5.04	
165	G	110.45	7.69	44.17					
166	E	117.87	8.13	54.32	29.34				175.22
167	G	110.18	8.49	43.99		110.43	8.54	4.48	172.74
168	L	120.82	7.65	51.73	43.45	120.77	7.65	3.80	
169	P			65.44	31.38				
170	G	105.33	8.96	45.64					
171	M	113.64	7.94	56.16	33.11				176.17
172	C	116.86	8.02	61.60	28.52	116.61	8.04	4.61	
173	G	112.95	9.41	44.81					172.62
174	G	110.20	8.27	44.55		110.41	8.31	4.33	171.59
175	A	118.93	8.07	49.79	23.11	119.03	8.09	5.29	174.60
176	L	125.20	7.95	52.89		125.47	7.97	4.63	
177	V	130.05	9.41	58.03	28.08				
178	S	117.92	8.69	61.12	62.36				
179	S								
180	N								
181	Q								
182	S				61.12				
183	I	110.44	6.50	59.95	35.44				
184	Q	116.60	7.83	56.58	24.95				175.14
185	N	109.79	7.99	55.50	36.87	109.69	7.96	3.79	171.34
186	A	110.04	5.79	52.21	19.04	110.53	5.86		
187	I	118.85	8.71	61.49					
188	L	123.49	7.49	55.60	43.98				178.90
189	G	102.05	7.32	45.32		102.44	7.35	4.14	171.14
190	I	118.40	7.91	55.35	39.87	118.46	7.92	5.25	
191	H	127.19	8.85	58.89	32.33				
192	V	119.03	8.78	61.34	34.93				
193	A	119.55	6.96	52.30	21.21				175.18
194	G	107.32	7.92	44.88		107.60	7.92	4.23	
195	G	96.26	7.23	44.54					
196	N	120.63	9.20	54.32	36.97				
197	S	108.12	8.65	59.75	62.11				
198	I	120.25	7.78	59.16	39.98				
199	L	124.46	8.28	53.12	41.87				
200	V	123.98	8.90	60.58	33.64				173.59
201	A	127.08	8.66	49.33	21.12	127.61	8.69	5.27	
202	K	124.0	8.70	54.27	29.74				176.27
203	L	128.24	8.40	56.55	41.62	128.81	8.41		175.82
204	V	129.08	7.93	60.94	34.73	129.27	7.97		
205	T	110.95	7.62	57.49	71.23				
206	Q	117.98	8.80	58.45	28.78				179.29
207	E	118.87	9.49	60.34	27.41	119.04	9.51	3.93	179.17
208	M	117.76	7.62	57.95	30.41	118.00	7.65	4.00	176.97
209	F	113.74	7.12	57.32	36.94	113.99	7.14	4.32	176.98
210	Q	119.31	7.47	57.92	27.84	119.65	7.48	4.09	
211	N	114.48	7.93	54.01	38.01	114.83	7.96	4.60	176.04
212	I	115.26	7.15	62.65	36.80	115.99	7.16	3.54	175.55
213	D	120.47	7.54	53.94	40.35	120.70	7.55	4.44	175.74

Residue	¹⁵ N	¹ H _N	¹³ C _α	¹³ C _β	¹⁵ N*	¹ H _N *	¹ H _α *	¹³ CO*
214 K	120.74	7.71	55.93	32.39	120.79	7.71	4.27	176.21
215 K	123.29	8.33	55.83	32.21	123.99	8.40	3.83	176.34
216 I	123.37	8.24	60.69	37.81	124.02	8.33	4.35	175.32
217 E	129.40	7.98	57.20	30.33	130.23	8.01	4.19	

Chemical shifts were directly (¹H) and indirectly (¹³C, ¹⁵N) referenced to DSS at 0.00 ppm as described by Wishart et al., (1995). * Chemical shifts at pH 7.0.

experiment sorted out many ambiguities that arose from problems of ¹³C_α chemical shift overlap. Using both ¹³C_α and ¹³C_β chemical shift information helped define and differentiate spin systems much better as random coil ¹³C_β chemical shifts for histidine, leucine, methionine, and glutamine are 29.0, 41.7, 32.4 and 29.2 ppm respectively. This ¹³C_β information was particularly useful when trying to identify what residue type was responsible for a given spin system. It also provided an additional data point when building chemical shift clusters.

The amide peak in the ¹H-¹⁵N HSQC spectrum that displayed the greatest intensity (at 129.40/7.98 ppm) corresponded to the C-terminus (Glu217). This intense peak was a favorable starting point and connectivities, in the 'i-1' direction, were made using information contained in HNCA, HN(CO)CA and HN(CA)CB spectra. This initial stretch continued until Ala193 where the resonance intensity in all spectra began to drop off. Resonance assignments (A193, G194, G195, I198, L199, etc.) were confirmed with the HSQC spectra collected from ¹⁵N residue-specific labelled HAV 3C protease. This integrated approach to sequentially assigning the backbone was repeated throughout the assignment process. Other starting points were Lys11 and Val109, as they belonged to helical regions that had been assigned using the initial ¹⁵N-edited experiments. Sequential stretches were assigned from these starting points and confirmed with residue-specific HSQC data. Once these sequentially assigned stretches could no longer be extended, other peaks were selected from the residue-specific HSQC spectra to serve as new starting points. Generally, the best check for correctness of a sequentially assigned cluster was to simply note the difference in ¹³C_α and ¹³C_β chemical shift

values between residues 'i' and 'i-1'. A correctly assigned cluster would rarely show chemical shift difference greater than 0.1 ppm.

Table 4.2 displays a couple regions (32-43, 87-96, 127-130 and 179-181) that could not be assigned. Generally, the residues that have been assigned on either side of these regions displayed very weak or no $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts and it is likely that those residues within the unassigned regions also have weak resonances. The weak resonance intensity could be due to the fact those residues are in mobile regions in some kind of intermediate or slow exchange process.

'Observed' versus 'predicted' HAV 3C protease chemical shifts. Comparisons between the observed and predicted chemical shifts for HAV 3C protease are displayed in Figure 4.11 (p.115). Predicted chemical shifts for ^{15}N , $^1\text{H}_\text{N}$, $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$ and ^{13}CO shifts were obtained from SHIFTX, whereas the predicted $^1\text{H}_\alpha$ chemical shifts were calculated using TOTAL (Williamson et al., 1995). Correlation coefficients of 0.76, 0.60, 0.97, 0.99, 0.78 and 0.78 were determined for ^{15}N , $^1\text{H}_\text{N}$, $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, $^1\text{H}_\alpha$ and ^{13}CO chemical shifts respectively. Although all these values indicate a 'good' correlation between observed and predicted data, the $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ correlations are nearly perfect. It is likely that these chemical shifts, relative to their chemical environment and structure, are more easily predicted than the other types of chemical shifts.

Correlation coefficients for ^{15}N , $^1\text{H}_\text{N}$, $^1\text{H}_\alpha$ and ^{13}CO are lower than those noted for the $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$. For this reason, we studied the eight residues that showed the greatest difference between observed and predicted ^{15}N , $^1\text{H}_\text{N}$, $^1\text{H}_\alpha$ and ^{13}CO chemical shifts (Table 4.3, p.116). This was done to determine if any trends could be found that might indicate structural differences between the solution and X-ray structures of HAV 3C protease. Trends that can be noted in Table 4.3 are I212 appearing in 3 (^{15}N , $^1\text{H}_\text{N}$, and ^{13}CO) of the 4 categories and 50 percent of the residues showing the greatest difference in ^{13}CO chemical shifts are glycines. I212

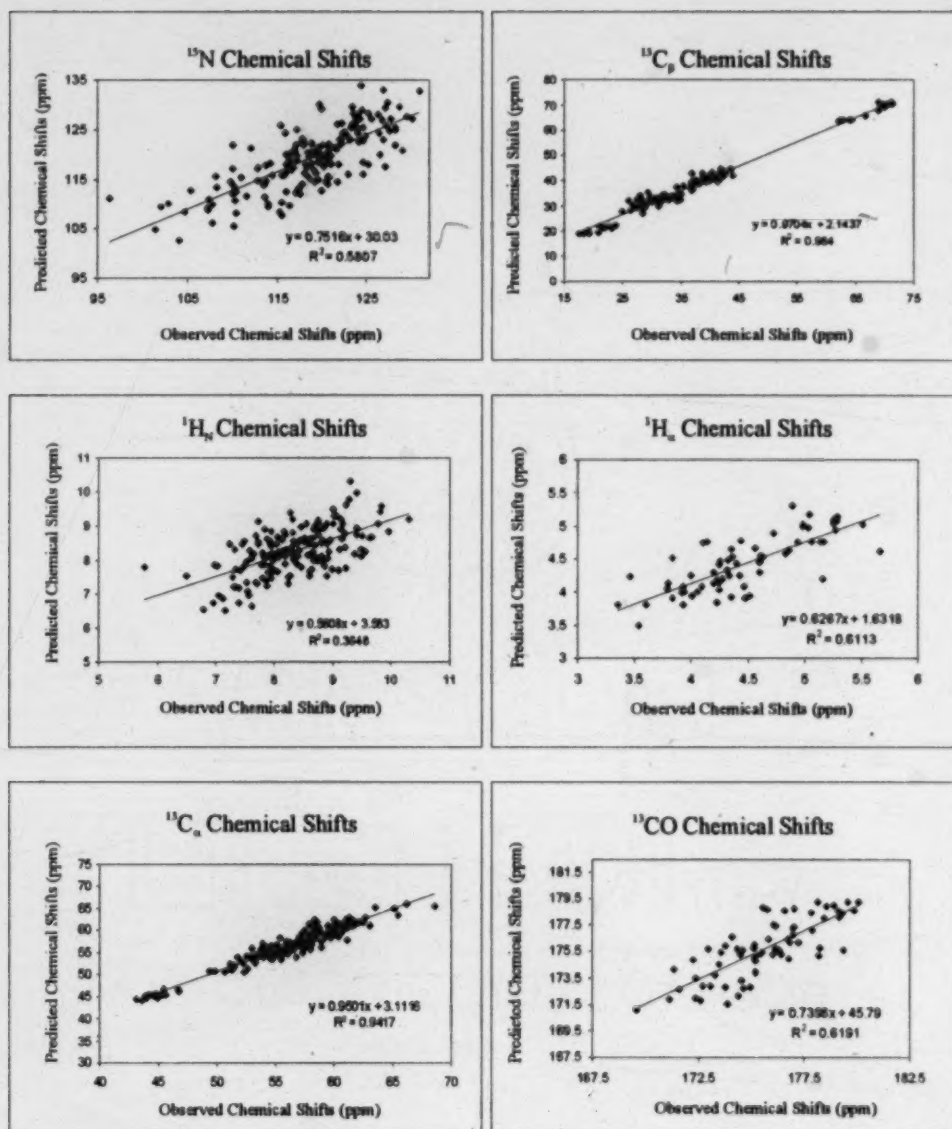


Figure 4.11. Predicted versus observed chemical shifts for HAV 3C protease. Predicted chemical shifts for ^{15}N , $^1\text{H}_n$, $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$ and ^{13}CO shifts were obtained from SHIFTX, whereas the $^1\text{H}_\alpha$ predicted chemical shifts came from the TOTAL program (Williamson et al., 1992).

Table 4.3. The 8 residues showing the greatest differences between observed and predicted ^{15}N , $^1\text{H}_\text{N}$, $^1\text{H}_\alpha$ and ^{13}CO chemical shifts.

^{15}N			$^1\text{H}_\text{N}$		
Residue	Observed	Predicted	Residue	Observed	Predicted
N114	115.46	125.74	H44	9.23	7.75
N124	127.05	117.60	F59	9.30	10.34
L135	119.73	129.93	Y66	9.40	9.96
A141	124.36	133.84	L155	8.90	7.55
Q159	112.02	121.23	A186	5.79	7.78
A186	110.04	121.80	L199	8.28	9.38
G194	96.26	111.14	M208	7.62	6.63
I212	115.26	108.29	I212	7.15	6.51
$^1\text{H}_\alpha$			^{13}CO		
Residue	Observed	Predicted	Residue	Observed	Predicted
V25	3.46	4.23	G62	174.95	172.80
R111	5.16	4.21	G63	173.87	171.60
L113	4.10	4.74	V85	178.20	175.21
L119	5.66	4.61	K105	179.34	175.61
D154	4.89	5.30	G125	174.41	172.10
G167	4.48	3.91	G173	172.62	171.80
G189	4.14	4.77	I212	175.55	178.84
K215	3.83	4.51	D213	175.74	178.69

was assigned using all of our spectra (^{15}N -edited and triple resonance) therefore it is unlikely that it is assignment incorrectly. This result may be indicating that the environment surrounding I212 is different in the solution-state compared to that in the solid-state. The result that 50 percent of the residues showing greatest differences in ^{13}CO chemical shifts are glycines may indicate that glycine ^{13}CO chemical shifts are more difficult to predict than other residue types. Otherwise,

those residues that are present in Table 4.3 do not show any obvious trends. As a result, it is very likely that there is no indication of structural differences, between solution and solid states of HAV 3C protease, that can be inferred from the differences found between our observed and predicted chemical shift sets.

Secondary structure of HAV 3C protease. The $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, $^1\text{H}_\alpha$ and ^{13}CO chemical shifts are sensitive to secondary structure. In fact, the secondary structure can be accurately predicted from these chemical shifts using the CSI method (Wishart et al., 1992, 1994). The CSI program (Wishart et al., 1994) was used to calculate indices for most amino acid residues in HAV 3C protease using available chemical shift information. Consequently, some stretches of secondary structure were predicted based on information from only one type of chemical shift and certain stretches (32-43, 87-96, 127-130 and 179-181) had no chemical shift assignments therefore no structure could be predicted. Obviously, a more accurate prediction of secondary structure could be attained as soon as all chemical shifts have been assigned. However, Table 4.4 displays the results of the CSI predictions based on the current set of chemical shift assignments. As can be seen, 11 β -strands and 2 α -helices have been predicted. The position and extent of these structural elements are in good agreement with structural elements in HAV 3C proteases crystal structure (Bergmann et al., 1997), as calculated by VADAR. VADAR (Wishart et al., 1994) used the 3D coordinates of the HAV 3C protease PDB file (1HAV) to calculate the secondary structure elements. It is likely that the differences between the CSI predictions and secondary structure from X-ray data would be significantly less if complete chemical shift assignments were present.

Table 4.4. Comparison of secondary structure features of HAV 3C protease obtained from CSI and X-ray diffraction methods. VADAR (Wishart, 1994) was used to calculate secondary structure from the PDB file (1HAV) of the X-ray crystal structure of HAV 3C protease (Bergmann et al., 1997). *From the PDB file (1HAV) as this α -helix was not calculated by VADAR.

β -Strands				α -Helices					
CSI				Vadar	CSI				Vadar
C $_{\alpha}$	C $_{\beta}$	H $_{\alpha}$	CO		C $_{\alpha}$	C $_{\beta}$	H $_{\alpha}$	CO	
				12-18	6-11		6-11	2-10	2-11
24-27				23-25					
				28-33					
				37-41					
58-61				55-60					
64-68	66-68			65-69					
72-76	73-76	73-76		73-76					
				84-90					
				95-97					
								105-108	108-111
				103-105					
115-123	121-123			115-122					
140-147	140-143			127-146					
149-157	151-153	151-155	150-154	152-158					
160-168	160-168			160-165					
175-177				174-179					
				186-188					
	191-194			190-193					
198-202				197-204					
					206-209		208-215	206-212	*206-213

4.5 Conclusion

In conclusion, 175 of the 210 (217 residues – 7 prolines) sequential backbone assignments have been made although a majority of the ^1H and ^{13}C assignments are still missing due to ^1H - ^{15}N TOCSY-HSQC, ^1H - ^{15}N NOESY-HSQC and HNCO experiments being performed on a HAV 3C protease sample at high pH. The three key experiments that enabled the near complete backbone assignments to be made were the “deuterium modified” HNCA, HN(CO)CA and the HN(CA)CB experiments as they yielded spectra with nearly all expected resonances. There were two reasons why these experiments were so successful. The first reason lies in the fact that we prepared this HAV 3C protease sample in solution conditions that minimized aggregation. The second reason was incorporating high levels of deuterium into the protease to combat the efficient relaxation rates of nuclei present in non-deuterated samples. The sequential

assignments that have been made are in good agreement with those chemical shifts that have been predicted by SHIFTX and TOTAL software packages based on the X-ray crystal structure of the HAV 3C protease. Finally, based on the 'Chemical Shift Index' and our present sequential assignments, secondary structure elements have been predicted that are in relatively good agreement with those structural elements determined by X-ray crystallography.

CHAPTER 5

General Discussion and Conclusion

This thesis centers on the 3C protease from Hepatitis A virus, the causal agent of one of the most common forms of acute, infectious hepatitis in the world. HAV 3C protease is a large, 217 residue protein that could serve as an excellent target for 'structure based drug design'. It is an excellent drug target because it is responsible for processing the polyprotein produced by the translation of viral mRNA and therefore, it is essential to the propagation of this virus (Linemeyer et al., 1985; Najarian et al., 1985; Palmenberg, 1990; Rueckert, 1996). The intent of this thesis was to obtain near complete ^1H , ^{13}C and ^{15}N backbone chemical shift assignments for HAV 3C protease. More specifically, in Chapter 2, I describe how NMR spectroscopy and ultracentrifugation studies were used in consort to identify solution conditions that minimized the macromolecular aggregation noted in our early HAV 3C protease NMR samples. In Chapter 3, I describe four expression and purification protocols that facilitated uniform ($^2\text{H}/^{13}\text{C}/^{15}\text{N}$, $^{13}\text{C}/^{15}\text{N}$, ^{15}N) and residue-specific (^{15}N) isotope incorporation into HAV 3C protease. Finally, in Chapter 4, I describe the heteronuclear NMR experiments, performed with single (^{15}N), doubly ($^{13}\text{C}/^{15}\text{N}$) and triply ($^2\text{H}/^{13}\text{C}/^{15}\text{N}$) labelled HAV 3C protease, that enabled 175 of the 217 backbone chemical shift assignments to be made. It is my contention that the sequential assignments that have been completed to date will enable NMR spectroscopy to be used in future 'drug-screening' efforts, as described in Shuker et al., (1996), to assist in the 'structure based drug design' of a protease inhibitor to fight Hepatitis A.

As I described previously in Chapters 1 and 2, to obtain "good" NMR spectra from larger proteins, it is important to ensure that the protein be monodisperse at concentrations of 1 mM or higher. Macromolecular aggregation is often problematic in protein NMR spectroscopy as it leads to extremely efficient ^{13}C , ^{15}N and ^1H T_2 relaxation rates which in turn leads to significantly broadened

lines, poor resolution and loss of signal. Early NMR experiments (^1H - ^{15}N TOCSY-HSQC, ^1H - ^{15}N NOESY-HSQC, HNCACB, CBCA(CO)NNH) performed with HAV 3C protease in 'high salt/high pH' conditions (350 mM NaCl, 20 mM KH_2PO_4 , 0.5 mM EDTA at pH 7.0) generally yielded spectra that were devoid of much of the expected chemical shift information. At that time, it was hypothesized that HAV 3C protease was behaving as an aggregate as efficient T_2 relaxation rates of nuclei in an aggregated complex could account for the significant loss of signal in these initial spectra. Additionally, $^3J_{\text{HNNH}\alpha}$ coupling constants calculated from ^1H linewidths in a HAV 3C protease HMQC-J spectrum using the method of Wishart and Wang, (1998) indicated that the molecule was behaving as a dimer. Sedimentation analysis and ^{15}N T_1 - and T_2 -envelope studies eventually confirmed this hypothesis.

To identify solution conditions that minimized HAV 3C protease aggregation, a series of 1D ^1H NMR spectra were collected with the protease dissolved in approximately 25 different buffer conditions. Buffer variables that were tested included pH, temperature, NaCl concentration and the addition of CHAPS, TFE, glycine and CaCl_2 . The quality of the amide region (sharpness and intensity of peaks) and half-height linewidth ($\Delta\nu_{1/2}$) of the Val25 γCH_3 resonance at -0.7 ppm were used to compare and contrast the behavior of HAV 3C protease under various buffer conditions. This was done because qualitative information about the presence of aggregation can be obtained by monitoring the broadening of NMR linewidths (Akiyoshi et al., 1993). Buffer conditions yielding the sharpest peaks in the amide region and the narrowest half-height linewidth for the Val25 γCH_3 resonance were 100 mM NaCl, 20 mM KH_2PO_4 , 0.5 mM EDTA and 1.0 mM DTT at pH 5.4. Subsequent sedimentation analysis, ^1H - ^{15}N HMQC-J and ^{15}N - T_1 and $-T_2$ envelope studies confirmed that HAV 3C protease was behaving as a monomer under these optimized conditions.

Chapter 2 also describes an important 'time-course' stability study of HAV 3C protease that involved collecting 1D ^1H NMR spectra of the protein over a

period of one month. This study determined that HAV 3C protease stock should be stored for up to a month at 4 °C in the post-column elution buffer (350 mM NaCl, 20 mM KH₂PO₄, 0.5 mM EDTA, 1.0 mM DTT at pH 5.4) under argon gas with 25 mL of 3 % (w/v) sodium azide added to ensure sterility. This study also determined that NMR experiments performed with HAV 3C protease yield the best results within 9 days of making the NMR sample. After the 9 day period, it is likely that some degree of HAV 3C protease aggregation occurs although the extent of this is unknown.

Chapter 3 describes the protocols developed to isotopically label HAV 3C protease with combinations of ²H, ¹³C and ¹⁵N. These protocols were based on the initial expression and purification of unlabelled HAV 3C protease that was published by Malcolm et al., (1993). Single (¹⁵N) and double (¹³C/¹⁵N) labelled HAV 3C protease was produced by substituting ¹⁵NH₄Cl and ¹³C₆-glucose for ¹⁴NH₄Cl and ¹²C-glucose into the media preparations. Obtaining a triple (²H/¹³C/¹⁵N) labelled HAV 3C protease required that our E. coli expression system be adapted to deuterium-based minimal media. This process entailed successively growing the bacteria in minimal media prepared with increasing amounts (50, 80 and 85 percent) of ²H₂O. Bacteria do not metabolize ²H₂O as efficiently as ¹H₂O due to thermodynamic and kinetic isotope effects (Galimov, 1985) therefore growth rates in ²H₂O based minimal media were much slower than those seen in pure ¹H₂O based minimal media. The uniform labelling strategies allowed isotopic incorporation levels of ≥ 95% for ¹³C and ¹⁵N and ~ 70 % for ²H.

To obtain ¹⁵N labelled HAV 3C protease with specific amino acids (alanine, glycine, leucine, lysine, isoleucine, phenylalanine and valine), we chose to use our prototrophic strain of E. coli and substitute an amino acid mixture (19 unlabelled and a single ¹⁵N labelled amino acid) for ¹⁴NH₄Cl in the minimal media (Muchmore et al., 1989). This simple residue-specific labelling protocol yielded good results with 5 of our 7 trials as only glycine and phenylalanine showed significant amounts of ¹⁵N 'scrambling'.

Chapter 4 reports the near complete (~85%) backbone sequential assignments of HAV 3C protease and the NMR experiments that were used to obtain them. Progress on the sequential assignments was greatly accelerated after we identified the 'low-salt/low-pH' solution conditions that minimized aggregation and after we prepared a triple ($^2\text{H}/^{13}\text{C}/^{15}\text{N}$) labelled HAV 3C protease to reduce efficient T_2 relaxation rates of ^1H , ^{13}C and ^{15}N nuclei. After preparing a 1 mM $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ labelled HAV 3C protease sample in the optimized 'low salt/low pH' solution conditions, "deuterium modified" CT-HNCA, CT-HN(CO)CA and HN(CA)CB experiments, designed in Dr. Lewis Kay's laboratory (Gardner and Kay, 1998; Yamazaki et al., 1994b), were completed that yielded spectra with nearly all of the expected resonances present. These spectra were used in consort with data from six (alanine, glycine, leucine, lysine, isoleucine and valine) ^{15}N residue-specific HSQC spectra to make the sequential backbone assignments.

Presently, backbone sequential assignments ($^1\text{H}_\text{N}$, ^{15}N , $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, ^{13}CO and $^1\text{H}_\alpha$) have been made for 175 of the 217 residues. ^{13}CO and ^1H resonances were acquired from HNCO, ^1H - ^{15}N TOCSY-HSQC and ^1H - ^{15}N NOESY-HSQC spectra collected at high pH. To determine ^{13}CO and $^1\text{H}_\alpha$ chemical shifts it was necessary to equate amide chemical shifts collected at low pH with those collected at high pH. For many amide resonances this relationship could not be unambiguously established, therefore a majority of these shifts are missing from the table of reported backbone chemical shifts in Chapter 4.

In Chapter 4 we also show that the backbone assignments exhibit good agreement with a set of predicted chemical shifts that had been estimated by SHIFTX and TOTAL (Williamson et al., 1995) software packages. These programs predicted ^1H , ^{13}C and ^{15}N chemical shifts based on the PDB file (1HAV) of the X-ray crystal structure of HAV 3C protease (Bergmann et al., 1997). In addition, the backbone chemical shifts were analyzed by the CSI program (Wishart

et al., 1994) demonstrating 2 α -helices and 11 β -strands. These structural elements are in good agreement with X-ray crystallographic data.

The efforts to obtain the sequential assignments of HAV 3C protease also served to teach an important lesson regarding NMR spectroscopy. Many NMR experiments were performed on HAV 3C protease prior to determining the optimum buffer conditions. Under these suboptimal conditions HAV 3C protease molecules had a tendency to aggregate leading to efficient T_2 relaxation rates and poor quality spectral data. However, significant progress occurred once a triply ($^2\text{H}/^{13}\text{C}/^{15}\text{N}$) labelled HAV 3C protease was prepared and placed in more optimal 'low salt/low pH' conditions. This optimized HAV 3C protease sample addressed the two key problems encountered in this project: 1) aggregation and 2) efficient relaxation rates. Essentially, an important take home message is that taking time to optimize your sample prior to acquiring important NMR spectra is a worthwhile endeavor for any protein study via NMR spectroscopy.

The completion of the sequential backbone assignments of HAV 3C protease will require some time and effort with our present collection of NMR spectra. However, to fill in the ^1H and ^{13}CO chemical shifts it will be necessary to perform other experiments. In order to determine all ^{13}CO chemical shifts, a CT-HNCO spectrum should be collected from a triply labelled HAV 3C protease sample under 'low salt-low pH' conditions. To determine all ^1H chemical shifts will likely require us to prepare a doubly labelled ($^{13}\text{C}/^{15}\text{N}$) HAV 3C protease sample and then collect a HCCH-TOCSY. It may even be advisable to prepare a new triply labelled protease sample for this last experiment to minimize the $^{13}\text{C}_\alpha$ T_2 relaxation rates. This new triply labelled protease should aim for 30-50% levels of deuterium incorporation (to ensure there are protons to be detected) therefore minimal media should be prepared with approximately 50 percent $^2\text{H}_2\text{O}$.

Finally, ligand-binding studies should be initiated using the "SAR by NMR" technique described by Shuker et al., 1996. The first ligand to be tested

should be iodoacetamide (an irreversible inhibitor) followed by the tightest binding inhibitor (as determined by enzymatic assay) that has been synthesized by Dr. John Vederas' group. These two ligands should cause perturbations in amide resonances that indicate binding to active site residues and possibly residues near the active site. This would provide a standard by which to measure other ligands. If the results from this step look good, it would be wise to test all ligands that have been screened enzymatically to determine if there are patterns of chemical shift perturbations that can be noted from inhibitors that bind with varying affinity. Information garnered from these experiments may allow us to hypothesize how the compounds are fitting into the protease and these hypotheses may be of utility to those synthesizing the compounds. Eventually, the solution structure of HAV 3C protease should be determined. Having structural information may allow one to use NMR to determine exactly how the ligands fit into the protease, information that has eluded crystallographers. This type of information would be significantly more attractive to the pharmaceutical companies.

Appendix A

Processing NMR spectra with NMRPIPE software

The following processing guidelines are based on the NMRPIPE software described by Delaglio et al., (1995). Steps that are typically involved in a NMR data processing session can be summarized in the following steps: (1) convert the FID from Varian to nmrPipe format, (2) use available scripts to process the data (this involves editing the processing shell script) and (3) inspect the resulting spectrum to decide if any additional processing is required such as re-phasing, zero-filling, or baseline correction. The NMRDRAW program can be used to view the processed spectrum.

Many processing functions and their associated arguments can be displayed by using function-specific commands. For example the following command will display information regarding the sine window function.

```
nmrPipe -fn SP -help
```

Following is a list of some of the functions that are typically used in data processing.

EXT	Extracts a region from the current dimension with specified limits.
FT	Applies a real or complex forward or inverse Fourier transform.
HT	Applies a Hilbert transform to reconstruct imaginary data.
LP	Linear prediction.
POLY	When used in frequency-domain, applies polynomial baseline correction. When used in the time-domain, applies solvent correction.
PS	Applies the zero- and first-order phase corrections.
REV	Reverse data order in given dimension
SOL	Applies solvent correction
SP	Applies sine-bell apodisation
ZF	Applies zero-filling

The following arguments are used by more than one function in the shell scripts above.

-di	Deletes imaginary data after the given processing function is performed.
-hdr	Extracts parameters recorded during previous processing from spectral header.
-in	Specifies the input file or file template.
-inPlace	Specifies replacement of the input data by the output result.
-inv	Activates the inverse mode of a given function.
-out	Specifies the output file or file template.
-ov	Permits overwriting of any pre-existing files.
-sw	Updates the spectral width and other ppm calibration information.
-verb	Permits processing in verbose mode, with status messages.

I Processing 2D NMR experiments.

CALIB. Copy the calib script into the directory containing the Varian propar. Run the calib script by simply typing calib. This invokes the script to search through the propar and find the various parameters required to transform the Varian data into nmripe data and store the parameters in a file called pipe.param. An example of a pipe.param file follows.

PARAMETERS FOR VAR2PIPE (NMRPIPE)

xN	1024	yN	128
xT	512	yT	64
xsw	6000.150	ysw	1400.000

1H OBServe frequency : 499.886
1H CARrier position : 4.739

15N OBServe frequency : 50.659
15N CARrier position : 118.504

FID.COM. Insert parameters listed in pipe.param and the appropriate path to the Varian 'FID' into the fid.com script using an editing program of your choice (eg. nedit). Invoke the script by typing fid.com. When this script is run it will take the fid (Varian FID) as input and transform it into nmripe format (eg. 15NHSQC.fid). A typical fid.com script follows.

```
var2pipe -in fid \
  -xT      512  -yT      64      \
  -xMODE    Complex -yMODE    Complex \
  -xSW      6000.150 -ySW      1400.000 \
  -xOBS      499.886 -yOBS      50.659 \
  -xCAR      4.739  -yCAR      118.504 \
  -xLAB      HN     -yLAB      N      \
  -ndim      2     -aq2D      States \
| nmrPipe -fn MAC -macro SNMRTXT/ranceY.M -noRd -noWr -out 15NHSQC.fid -verb -ov
```

FT2D.COM. The ft2d.com script will transform '15NHSQC.fid' into the '15NHSQC.ft2' which is in a format that can be viewed by nmrDraw. It is important to know what portion of the spectra (i.e. the amide region) that requires processing. Insert this defined region into the EXT or extract function of this script. A typical ft2d.com script follows.

```
#!/bin/csh
nmrPipe -in 15NHSQC.fid \
| nmrPipe -fn POLY -time \
| nmrPipe -fn SP -off 0.33 -end 0.98 -pow 1 -c 1.0 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT \
| nmrPipe -fn PS -p0 8.0 -p1 -35.0 \
| nmrPipe -fn EXT -x1 12.5PPM -xn 5.5PPM -sw -verb -di \
| nmrPipe -fn TP \
| nmrPipe -fn LPC -fb -auto \
| nmrPipe -fn SP -off 0.417 -end 0.98 -pow 2 -c 1.0 \
| nmrPipe -fn ZF -auto -verb \
| nmrPipe -fn FT \
| nmrPipe -fn PS -p0 303.0 -p1 133.0 -di \
| nmrPipe -fn POLY -auto \
| nmrPipe -fn TP \
| nmrPipe -fn POLY -auto \
-verb -ov -out 15NHSQC.ft2
```


You will see this script has two 'PS' lines. The first is for phasing in the X dimension ($^1\text{H}_\text{N}$) and the second is for phasing in the Y dimension (^{15}N).

If the spectrum is out of phase in the X or Y dimension, phase the spectrum appropriately in nmrDraw with p0 and p1 slide bars. Note the numbers associated with your phasing in nmrDraw and insert them (taking into consideration what their previous values) into the ft2d.com script. Rerun your ft2d.com script.

II Processing 3D NMR experiments

CALIB. Run the calib script in the directory containing the 3D Varian propar. Invoke by typing calib. Ensure that both ^{15}N and ^{13}C information is collected when asked by the script. This searches the propar file to find the parameters that are necessary to convert the Varian FID to nmripe format. A typical pipe.param file for 3D experiments follows.

PARAMETERS FOR VAR2PIPE (NMRIPIPE)

xN	1024	yN	100	zN	64
xT	512	yT	50	xT	32
xsw	6000.150	ysw	7794.232	zsw	1600.000

1H OBServe frequency : 499.886
1H CARrier position : 4.739

15N OBServe frequency : 50.659
15N CARrier position : 117.821

13C OBServe frequency : 125.701
13C CARrier position : 43.016

FID_EXPERIMENT.COM. Insert parameters into the 'fid_experiment.com' script using an editor of your choice. Run the script by typing fid_experiment.com. The input is the varian 'fid' which will be transformed and stored into several nmrpipe format 'fids' in a directory called FID. Following is a typical fid_experiment.com script.

```
#!/bin/csh
var2pipe -in fid \
-xN 1024 -yN 100 -zN 64 \
-xT 512 -yT 50 -zT 32 \
-xMODE Complex -yMODE Complex -zMODE Complex \
-xSW 6000.150 -ySW 7794.232 -zSW 1600.0 \
-xOBS 499.886 -yOBS 125.701 -zOBS 50.659 \
-xCAR 4.739 -yCAR 43.016 -zCAR 117.821 \
-xLAB H -yLAB C -zLAB N \
-ndim 3 -aq2D STATES -aqORD 0 \
#| nmrPipe -fn MAC -macro SNMRTXT/ranceZ.M -noRd -noWr \
| pipe2xyz -out FID/experiment%03d.fid -verb -ov
```

FT2D_EXPERIMENT.COM. Processing in 3 dimensions takes a very long time. It is a good idea to observe your data in 2 dimensions prior to processing the experiment fully in 3 dimensions. For this use the ft2d_experiment.com script. This allows you to view the full experiment in one xy plane. Invoke the script as before.

```
#!/bin/csh
nmrPipe -in FID/experiment001.fid \
| nmrPipe -fn SOL \
| nmrPipe -fn SP -off 0.33 -end 0.99 -pow 1 -c 1.0 \
| nmrPipe -fn ZF -size 512 \
| nmrPipe -fn FT -verb \
| nmrPipe -fn PS -p0 -0.0 -p1 0 -di \
| nmrPipe -fn EXT -x1 10.5PPM -xn 5.75PPM -sw \
| nmrPipe -fn TP \
| nmrPipe -fn LP -fb \
| nmrPipe -fn SP -off 0.43 -end 0.99 -pow 2 -c 1.0 \
| nmrPipe -fn ZF -size 512 \
| nmrPipe -fn FT -verb \
| nmrPipe -fn PS -p0 -90.0 -p1 180.0 -di \
| nmrPipe -fn TP \
| nmrPipe -fn POLY -auto \
-ov -out experiment.ft2
```

The 'experiment.ft2' (output) file can be viewed by nmrDraw. It is likely that this spectrum will have to be phased with the p0 and p1 toggle bars. Insert the appropriate values back into the ft2d_experiment.com script and rerun.

FT3D.COM. Initially it is important to transfer all the relevant values from the ft2d.com script to the ft3d.com script. Likely, this will just be the phasing values you have determined. Invoke the script to fully process your 3D experiment by typing ft3d.com. An example script follows.

```
#!/bin/csh
#
# Processing of F3 dimension (HN)
#
xyz2pipe -in FID/experiment%03d.fid -x -verb \
| nmrPipe -fn SOL \
| nmrPipe -fn SP -off 0.33 -end 0.99 -pow 1 -c 1.0 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT \
| nmrPipe -fn PS -p0 0.0 -p1 0.0 \
| nmrPipe -fn EXT -x1 5.75PPM -xn 10.5PPM -di -sw \
| pipe2xyz -out DAT/experiment%03d.DAT -x -ov
#
# Processing of F2 dimension (N) without LP
#
xyz2pipe -in DAT/experiment%03d.DAT -z -verb \
| nmrPipe -fn SP -off 0.5 -end 0.95 -pow 1 -c 1.0 \
| nmrPipe -fn ZF -size 72 \
| nmrPipe -fn FT -di \
| pipe2xyz -out DAT/experiment%03d.DAT -z -inPlace
#
# Processing of F1 dimension (C) with LP
#
xyz2pipe -in DAT/experiment%03d.DAT -y -verb \
| nmrPipe -fn LP -fb -ps90-180 \
| nmrPipe -fn SP -off 0.5 -end 0.99 -pow 2 -c 1.0 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT \
| nmrPipe -fn PS -p0 -90.0 -p1 180.0 -di \
| pipe2xyz -out DAT/experiment%03d.DAT -y -inPlace
#
# Processing of F2 dimension (N) with LP
#
xyz2pipe -in DAT/experiment%03d.DAT -z -verb \
| nmrPipe -fn HT \
| nmrPipe -fn FT -inv \
| nmrPipe -fn ZF -size 72 -inv \
| nmrPipe -fn SP -off 0.5 -end 0.95 -pow 1 -c 1.0 -inv \
| nmrPipe -fn LP -fb -ps0-0 \
| nmrPipe -fn SP -off 0.417 -end 0.98 -pow 2 -c 1.0 \
| nmrPipe -fn ZF -auto \
| nmrPipe -fn FT -di \
| pipe2xyz -out DAT-2/experiment%03d.DAT -z -inPlace
```

First run the `ft3d.com` commenting out the LP (linear prediction) and ZF (zero filling) commands which are very time intensive. Have a look at your final data from this initial processing. If things seem to be appropriate, run the script with LP and ZF. This will take a couple of hours. To be considerate to others use the following command.

```
nice +19 ft3d_experiment.com > & ft3d.log &
```

You can then periodically check the progress by viewing the `ft3d.log`.

References

- Abragam, A. *The Principles of Nuclear Magnetism*, Clarendon Press, Oxford, 1961. pp. 309.
- Akiyoshi, K., Deguchi, S., Moriguchi, N., Yamaguchi, S., and Sunamoto, J. (1993) *Macromolecules* 26:3062-3068.
- Allaire, M., Chernai, M.M., Malcolm, B.A., and James, M.N. (1994) *Nature* 369:72-76.
- Allerhand, A., Childers, R.F. and Oldfield, E., (1973) *Biochemistry* 12:1335-1341.
- Anglister, J., Grzesiek, S., Ren, H., Klee, C.B., and Bax, A. (1993) *Journal of Biomolecular NMR* 3:121-126.
- Bachovchin, W.M. (1985) *Proceedings of the National Academy of Sciences of the USA* 82:7948-7952.
- Bax, A. and Davis, D.G. (1985) *Journal of Magnetic Resonance* 65:355-360.
- Bax, A., Clore, G.M., Driscoll, P.C., Gronenborn, A.M., Ikura, M., and Kay, L.E. (1990) *Journal of Magnetic Resonance* 87:620-627.
- Bax, A., Sparks, S.W., and Torchia, D.A. (1989) *Methods in Enzymology* 176:134-150.
- Bergmann, E.M., and James, M.N.G. (1999) *Proteases of Infectious Agents: Proteolytic Enzymes of the Viruses of the Family Picornaviridae*. Academic Press.
- Bergmann, E.M., Mosimann, S.C., Chernaia, M.M., Malcolm, B.A., James, M.N., (1997) *Journal of Virology* 71(3):2436-2448.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi T., Tasumi, M., (1977) *European Journal of Biochemistry* 80(2):319-324.
- Billeter, M., Braun, W. and Wuthrich, K. (1982) *Journal of Molecular Biology* 155:321-346.
- Blundell, T. L. and Johnson, L. N. (1976) *Protein Crystallography*, Academic Press.
- Braunschweiler, L. and Ernst, R.R. (1983) *Journal of Magnetic Resonance* 53:521-528.

Brill, B.M., Kati, W.M., Montgomery, D., Karwowski, J.P., Humphrey, P.E., Jackson, M., Clement, J.J., Kadam, S., Chen, R.H., and McAlpine, J.B. (1997) *Journal of Antibiotics* 49:541-546.

Browne, D.T., Kenyon, G.L., Packer, E.L., Sternlicht, H., and Wilson, D.M. (1973) *Journal of the American Chemical Society* 95:1316-1323.

Bystrov, V.F. (1976) *Progress in NMR Spectroscopy* 10:41-81.

Carpenter, J.F. (1998) *Journal of Pharmaceutical Sciences* 12:1597-1603.

Cavanagh, J., Fairbrother, W.J., Palmer, A.G. III, Skelton, N.J. (1996) *Protein NMR Spectroscopy Principles and Practice*, Academic Press.

Center for Disease Control and Prevention (<http://www.cdc.gov/>).

Clore, G.M. and Gronenborn, A.M. (1987) *Protein Engineering* 1:275-288.

Clore, G.M., and Gronenborn, A.M. (1991) *Progress in NMR Spectroscopy* 23:43-92.

Clore, G.M. and Gronenborn, A.M. (1994) *Protein Science* 3:372-290.

de Dios, A.C., Pearson, J.G. and Oldfield, E., (1993) *Science* 260:1491-1495.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and A. Bax (1995) *Journal of Biomolecular NMR* 6:277-293.

Delaglio, F., Torchia, D.A. and Bax, A. (1991) *Journal of Biomolecular NMR* 1:439-446.

Derman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.H., Srinivasan, A.R., Schneider, B., (1992) *Biophysical Journal* 63(3):751-759.

DesJarlais, R.L., Seibel, G.L., Kuntz, I.D., Furth, P.S., Alvarez, J.C., Ortiz de Monteillano, P.R., Decamp, D.L., Babe, L.M., Craik, C.S., (1990) *Proceeding of the National Academy of Sciences of the USA* 87(17):6644-6648.

Ernst, R.R. and Anderson, W.A., (1966) *Review of Scientific Instruments* 37:93-102.

Farrow, N.A., Zhang, O., Forman-Kay, J.D. and Kay, L.E. (1995) *Biochemistry* 34 (3):868-878.

Fesik, S.W. (1993) *Journal of Biomolecular NMR* 3:261-269.

Fesik, S.W. and Zuiderweg, E.R.P. (1988) *Journal of Magnetic Resonance* 78:588-593.

Frey, M.H., Wagner, G., Vasak, M., Sorensen, O.W., Neuhaus, D., Worgotter, E., Kagi, J.H.R., Ernst, R.R. and Wuthrich, K.J. (1985) *Journal of the American Chemical Society* 107:6847-6851.

Galimov, E.M. (1985) *The Biological Fractionation of Isotopes*, Orlando, Academic, p. 261.

Gardner, K.H. and Kay, L.E. (1998) *Annual Review of Biophysics and Biomolecular Structure* 27:357-406.

Garrett, D.S., Poweres, R., Gronenborn, A.M. and Clore, G.M. (1991) *Journal of Magnetic Resonance* 95:214-220.

Goody, R.S. (1995) *Nature Medicine* 1(6):519-520.

Gorbalenya, A.E. and Snijder, E.J. (1996) *Viral cysteine proteinases. Perspectives Drug Discovery and Design* 6:64-86.

Graff, J., Richards, O.C., Swiderek, K.M., Davis, M.T., Rusnak, F., Harmon, S.A., Jia, X.J., Summers, D.F. and Ehrenfeld, E. (1999) *Journal of Virology* 73:6015-6023.

Grzesiek, S. and Bax, A. (1992) *Journal of Magnetic Resonance* 96:432-440.

Grzesiek, S. and Bax, A. (1993) *Journal of the American Chemical Society* 115:12593-12594.

Grzesiek, S., Aglister, J., Ren, H. and Bax, A. (1993) *Journal of the American Chemical Society* 115:4369-4370.

Hajduk, P.K., Meadows, R.P. and Fesik, S.W. (1997) *Science* 278:497-499.

Heinz, B.A., Tang, J., Labus, J.M., Chadwell, F.W., Kaldor, S.W., Hammond, M., (1996) *Antimicrobial Agents and Chemotherapy* 40(1):267-270.

Henry, G.D. and Sykes, B.D. (1992) *Biochemistry* 31(23):5284-5297.

Ikura, M., Kay, L.E., and Bax, A. (1990) *Biochemistry* 29:4659-4667

James, M.N.G. (1993) *Proteolysis and Protein Turnover: Convergence of active-centre geometries among the proteolytic enzymes*. Bond, J.S. and Barrett, (eds). Portland Press, London.

- Jeener, J., (Lecture, Ampere Summer School, Basko Polje, Yugoslavia, 1971)
- Jeener, J., Meier, B.H., Bachmann, P. and Ernst, R.R. (1979) *Journal of Chemical Physics* 71:4546-4550.
- Jewell, D.A., Swietnicki, W., Dunn, B.M., and Malcolm, B.A. (1992) *Biochemistry* 31:7862-7869.
- Jia, X.Y., Ehrenfeld, E., and Summers, D.F. (1991) *Journal of Virology* 65:2595-2600.
- Jia, X.Y., Summers, D.F. and Ehrenfeld, E. (1993) *Virology* 193:515-519.
- Johnson, M.L., Correia, J.J., Yphantis, D.A. and Halvorson, H.R. (1981) *Biophysical Journal* 36:575-588.
- Kaarsholm, N.C. and Ludvigsen, S. (1995) *Receptor* 5:1-8.
- Karplus, M. (1959) *Journal of Physical Chemistry* 30:11-15.
- Kay, L.E., Brooks, B., Sparks, S.W., Torchia, D.A. and Bax, A. (1989) *Journal of the American Chemical Society* 111:5488-5490.
- Kay, L.E. and Bax, A. (1990) *Journal of Magnetic Resonance* 86:110-126.
- Kay, L.E., Clore, G.M., Bax, A., and Gronenborn, A.M. (1990a) *Science* 249:411-414.
- Kay, L.E., Ikura, M. and Bax, A. (1990b) *Journal of the American Chemical Society* 112:888-889.
- Kay, L.E., Ikura, M., Tschudin, R., and Bax, A. (1990c) *Journal of Magnetic Resonance* 89:496-514.
- Kay, L.E., Keifer, P., and Saarinen, T. (1992) *Journal of the American Chemical Society* 114:10663-10665.
- Kay, L.E., Xu, G.Y. and Yamazaki, T. (1994) *Journal of Magnetic Resonance, Series A* 109:129-133.
- Kay, L.E. (1995) *Progress in Biophysics and Molecular Biology* 63:277-299.
- Kay, L.E. (1997) *Biochemistry and Cell Biology* 75:1-15.
- Kay, L.E. and Gardner, K.H. (1997) *Current Opinion in Structural Biology* 7:722-731.

- Kessler, H., Gehrke, M., Griesinger, C. (1988) *Angewandte Chemie International Edition in English* 27:490-536.
- Koff, R.S. (1998) *The Lancet* 351:1643-1648.
- Kreilgaard, L., Jones, L.S., Randolph, T.W., Frokjaer, S., Flink, J.M., Manning, M.C. and
- Kumar, A., Ernst, R.R. and Wuthrich, K. (1980) *Biochemical and Biophysical Research Communications* 95:1-15.
- Laemmli, U.K. (1970) *Nature* 227:680-685.
- LeMaster, D.M. and Richards, F.M. (1985) *Biochemistry* 24:7263-7270.
- Linemeyer, D.L., Menke, J.G., Martin-Gallardo, A., Huges, J.V., Young, A., and Mitra, S.W. (1985) *Journal of Virology* 54:347-355.
- London, R.E. (1990) *Journal of Magnetic Resonance* 86:410-415.
- MacLachlan, L.K., Middleton, D.A., Andrew, E.J. and Reid, D.G. (1997) *Methods in Molecular Biology, Vol. 60: Protein NMR Techniques*, Reid, D.G. (ed), Humana Press Inc., Totowa, NJ, pp. 337-362.
- Malcolm, B.A. (1995) *Protein Science* 4:1439-1445.
- Malcolm, B.A., Chin, S.M., Jewell, D.A., Stratton-Thomas, J.R., Thudium, K.B., Ralston, R. and Rosenberg, S. (1992) *Biochemistry* 31(13):3358-3363.
- Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A. and Bax, A. (1989) *Journal of the American Chemical Society* 111:1515-1517.
- Martin, A. (1993) *Physical Pharmacy, 4th Edition*. Lea and Febiger, Philadelphia, London, Chapter 10, 212-250.
- Martin, A., Benichou, D., Chao, S.F., Cohen, L.M. and Lemo, S.M. (1999) *Journal of Virology* 73:6220-6227.
- McDonald, C.C. and Phillips, W.D., (1967) *Journal of the American Chemical Society* 89:6332-6336.
- McIntosh, L.P., Griffey, R.H., Muchmore, D.C., Neilson, C.P., Redfield, A.G., and Dahlquist, F.W. (1987) *Proceedings of the National Academy of Sciences of the USA* 84:1244-1248.

McIntosh, L.P. and Dahlquist, F.W. (1990) *Quarterly Review of Biophysics* 23:1-38.

McIntosh, L.P., Wand, A.J., Lowry, D.F., Redfield, A.G. and Dahlquist, F.W. (1990) *Biochemistry* 29:6341-6362.

Montelione, G.T., Wuthrich, K., Burgess, A.W., Nice, E.C., Wagner, G., Gibson, K.D., and Scheraga, H.A. (1992) *Biochemistry* 31:236-249.

Moore, J.M. (1999) *Current Opinion in Biotechnology* 10:54-58.

Morris, G.A. and Freeman, R. (1979) *Journal of the American Chemical Society* 101:760-762.

Morris, T.S., Frommann, S., Shechosky, S., Lowe, C., Lall, M.S., Gauss-Muller, V., Purcell, R.H., Emerson, S.U., Vederas, J.C., Malcolm, B.A., (1997) *Bioorganic Medicinal Chemistry* 5(5):797-807.

Mossakoska, D.E., and Smith, R.A.G. (1997) *Methods in Molecular Biology, Vol. 60: Protein NMR Techniques*, Reid, D.G. (ed), Humana Press Inc., Totowa, NJ, pp. 325-335.

Muchmore, D.C., McIntosh, L.P., Russell, C.B., Anderson, D.E. and Dahlquist, F.W. (1989) *Methods in Enzymology* 177:44-73.

Muhandiram, D.R. and Kay, L.E. (1994) *Journal of Magnetic Resonance Series B* 103:203-216.

Najarian, R., Caput, D., Gee, W., Potter, S.J., Renard, A., Merryweather, J., Van Nest, G., and Dina, D. (1985) *Proceedings of the National Academy of Science of the USA* 82:2627-2631.

Neidhardt, F.C., (ED.) (1987) *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology, vol. I*. Washington, D.C: Amer. Soc. Microbiol.

NonLinear Least Squares Program for analysis of Equilibrium Ultracentrifugation Experiments. Copyright © 1991 by David A. Yphantis. 99 River Rd., Mansfield Center, Connecticut 06250-1018.

Olson, A.J. and Goodsell, D.S. (1998) *SAR QSAR Environmental Research* 8(3-4):273-285.

Optima XL-1 Analytical Ultracentrifuge Instruction Manual. Published by the Spinco Business Center of Beckman Instruments, Inc. Palo Alto, CA. (1997)

- Oschkinat, H., Griesinger, D., Kraulis, P.J., Sorensen, O.W., Ernst, R.R., Gronenborn, A.M., Clore, G.M. (1988) *Nature* 332:374-376.
- Palmenberg, A.C. (1990) *Annual Review of Microbiology* 44:602-623.
- Pardi, A. (1992) *Current Opinion in Structural Biology* 2:832-835.
- Pardi, A., Billeter, M. and Wuthrich, K. (1984) *Journal of Molecular Biology* 180:741-751.
- Pelton, J.G., and Wemmer, D.E., (1995) *Annual Review of Physical Chemistry* 46:139-167.
- Perona, J.J. and Craik, C.S. (1995) *Protein Science* 4:337-360.
- Petithory, J.R., Masiarz, F.R., Kirsch, J.F., Santi, D.V., and Malcolm, B.A. (1991) *Proceedings of the National Academy of Sciences of the USA* 88:11510-11514.
- Powers, J.C., Odake, S., Oleksyszyn, J., Hori, H., Ueda, T., Boduszek, B., Kam, C., (1993) *Agents and Actions Supplement* 42:3-18.
- Powers, R., Garrett, D.S., March, C.J., Frieden, E.A., Gronenborn, A.M. and Clore, G.M. (1992) *Biochemistry* 31:4334-4346.
- Rasnick, D. (1996) *Perspectives in Drug Discovery and Design* 6:47-63.
- Reid, D.G., MacLachlan, L.K., Edwards, A.J., Hubbard, J.A. and Sweeney, P.J. (1997) *Methods in Molecular Biology, Vol. 60: Protein NMR Techniques*, Reid, D.G. (ed), Humana Press Inc., Totowa, NJ, pp. 1-28.
- Richards, F.M. (1974) *Journal of Molecular Biology* 82:1-14.
- Richmond, T.J. (1984) *Journal of Molecular Biology* 178:63-89.
- Roberts, G.C.K. (1999) *Current Opinion in Biotechnology* 10:42-47.
- Roth, S.M., Schneider, D.M., Strobel, L.A., vanBerkum, M.B., Means, A.R. and Wand, A.J. (1992) *Biochemistry* 31:1443-1451.
- Rubin, B., Antonaccio, M.J. and Horovitz, Z.P. (1978) *Progress in Cardiovascular Disease* 21(3):183-194.

Rueckert, R.R. (1996) *Fields Virology: Picornaviridae: The Viruses and their Replication*. (Fields, B.N., Knipe, D.M., Howley, P.M., Channock, R.M., Melnick, J.L., Monath, T.P., Roizmann, B. and Straus, S.E. eds). Lippincott-Raven, Philadelphia.

Schechter, T. and Berger, A. (1967) *Biochemical and Biophysical Research Communications* 27:157-162.

Schein, C.H. (1990) *Bio/Technology* 8:308-315.

Schiffer, C.A., Huber, R., Wuthrich, K., van Gunsteren, W.F., (1994) *Journal of Molecular Biology* 241(4):588-599.

Shan, X., Gardner, K.H., Muhandiram, D.R., Kay, L.E., Arrowsmith, C.H., (1998) *Journal of Biomolecular NMR* 3:307-318.

Shuker, S.B., Hajduk, P.J., Meadows, R.P. and Fesik, S.W. (1996) *Science* 274:1531-1534.

Sjogren, M.H. (1998) *Hepatology* 27:887-888.

Skelton, J.J., Aspiras, F., Ogez, J., and Schall, T.J. (1995) *Biochemistry* 34(16):5329-5342.

Skelton, N.J., Aspiras, F., Ogez, J. and Schakk, T.J. (1995) *Biochemistry* 34:5329-5342.

Skiles, J.W., and McNeil, D. (1990) *Tetrahedron Letters* 31:7277-7280.

Spyracopoulos, L., Gagne, S.M., and Sykes, B.D., (1999) *Proceedings for the International School of Structural Biology and Magnetic Resonance. 4th Course on Dynamics, Structure and Function of Biological Macromolecules*

Thiel, T.K. (1998) *American Family Physician* 57:1500-1501.

Tjandra, N., Wingfield, P.T., Stahl, D.J., and Bax, A. (1996) *Journal of Biomolecular NMR* 8:273-284.

Venters, R.A., Huang, C.C., Ramer, B.T., Trolard, R., Spicer, L.D., and Fierke, C.A. (1995) *Journal of Biomolecular NMR* 5:339-344.

Vento, S., Garofano, T., Renzini, C., Cainelli, F., Casali, F., Ghironzi, G., Ferraro, T., and Conaia, E. (1998) *New England Journal of Medicine* 338:286-290.

Voet, D. and Voet, J.G. (1995) *Biochemistry, 2nd Edition*. John Wiley and Sons. New York. pp. 389-400.

- Wagner, G. and Wuthrich, K. (1982) *Journal of Molecular Biology* 155:347-366.
- Wagner, G., (1993) *Journal of Biomolecular NMR* 3:375-385.
- Wang, A.C. and Bax, A. (1995) *Journal of the American Chemical Society* 113:7772-7777.
- Wang, A.C. and Bax, A. (1996) *Journal of the American Chemical Society* 117:1810-1813.
- Wider, G. and Wuthrich, K. (1999) *Current Opinion in Structural Biology* 9(5):594-601.
- Williamson, M.P., Havel, T.F., Wuthrich, K., (1985) *Journal of Molecular Biology* 182(2):295-315.
- Williamson, M.P., Asakura, T., Nakamura, E. and Demura, M. (1992) *Journal of Biomolecular NMR* 1:83-98.
- Williamson, M.P., Kikuchi, J. and Asakura, T. (1995) *Journal of Molecular Biology* 247(4):541-546.
- Wishart, D.S., Sykes, B.D., Richards, F.M. (1991) *Journal of Molecular Biology* 222:311-333.
- Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) *Biochemistry* 31:1647-1651.
- Wishart, D.S. and Sykes, B.D. (1994a) *Journal of Biomolecular NMR* 4:171-180.
- Wishart, D.S. and Sykes, B.D. (1994b) *Methods in Enzymology* 39:363-392.
- Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. and Sykes, B.D. (1995a) *Journal of Biomolecular NMR* 5:67-81.
- Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995b) *Journal of Biomolecular NMR* 6:135-140.
- Wishart, D.S., Fortin, S., Woloschuk, D.R., Wong, W., Rosborough, T., Van Domselaar, G., Schaeffer, J. and Szafron, D. (1997) *CABIOS* 13(5):561-562.
- Wishart, D.S. and Wang Y. (1998) *Journal of Biomolecular NMR* 11:329-336.
- Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, New York, John Wiley & Sons.

Wuthrich, K., Wider, G., Wagner, G. and Braun, W. (1982) *Journal of Molecular Biology* 155:311-319.

Yamazaki, T., Lee, W., Revington, M., Mattiello, D.L., Dahlquist, F.W., Arrowsmiths, C.H. and Kay, L.E. (1994a) *Journal of the American Chemical Society* 116:6464-6465.

Yamazaki, T., Lee, W., Arrowsmith, C.H., Muhandiram, D.R. and Kay, L.E. (1994b) *Journal of the American Chemical Society* 116:11655-11666.

Zhang, O., Kay, L.E., Olivier, J.P. and Forman-Kay, J.D. (1994) *Journal of Biomolecular NMR* 4:845-858.